

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 18:15:11 ; Search time 18.37 Seconds  
(without alignments)  
1199.035 Million cell updates/sec

Title: US-09-405-504A-53

Perfect score: 3384

Sequence: 1 MLLGASLVGVLLFSKLVKL.....RYVPLDQAYSRIQAGEEKL 643

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query No.	Score	Match	Length	DB ID	Description
1	2110	62.4	646	1	FATP_MOUSE	Q60714 mus musculus
2	2087	61.7	646	1	FATP_RAT	P97849 rattus norv
3	1064	31.4	620	1	VLCS_HUMAN	O14975 homo sapien
4	1044	30.9	620	1	VLCS_RAT	P97524 rattus norv
5	1022	30.2	620	1	VLCS_MOUSE	O35488 mus musculus
6	714	21.1	623	1	FAT1_YEAST	P38225 saccharomyc
7	350.5	10.4	522	1	CAIC_ECOLI	P31552 escherichia
8	318	9.4	661	1	ACSA_COPCI	O13440 coprinus ci
9	311.5	9.2	561	1	LCFA_ECOLI	P29212 escherichia
10	303	9.0	545	1	4CL1_SOLTU	P31684 solanum tub
11	300	8.9	545	1	4CL2_SOLTU	P31685 solanum tub
12	299	8.8	547	1	4CL1_TORAC	O24145 nicotiana t
13	293.5	8.7	561	1	4CL3_ARATH	Q98777 arabidopsis
14	286	8.5	544	1	4CL1_PETCR	P14912 petroselinu
15	282	8.3	544	1	4CL2_PETCR	P14913 petroselinu
16	278.5	8.2	553	1	4CL_VANPL	O24540 vanilla pla
17	271.5	8.0	562	1	LCFA_HAEIN	P46450 haemophilus
18	271	8.0	543	1	FAT2_YEAST	P38137 saccharomyc
19	270.5	8.0	656	1	ACSA_RHOCA	O68040 rhodobacter
20	267.5	7.9	542	1	4CL2_TORAC	O24146 nicotiana t
21	262	7.7	694	1	ACSA_CRYPV	Q27549 cryptospori
22	260	7.7	5255	1	BACA_BACLI	O68006 b bacitraci
23	259	7.7	670	1	ACSA_EMENI	P16928 emeritella
24	259	7.7	3587	1	SRF1_BACSU	P27206 bacillus su
25	255.5	7.6	3587	1	SRF2_BACSU	Q04747 bacillus su
26	255	7.5	669	1	ACSA_PENCH	P36333 penicillium
27	253	7.5	6359	1	BACC_BACLI	O68008 b bacitraci
28	251	7.4	556	1	4CL2_ARATH	O9s725 arabidopsis
29	250	7.4	3587	1	TYCB_BACBR	O30408 b tyrocidin
30	249	7.4	548	1	YDID_ECOLI	P38135 escherichia
31	247	7.3	675	1	ACSA_CANAL	O94049 candida alb
32	246	7.3	713	1	ACSL_YEAST	O1574 saccharomyc
33	246	7.3	2560	1	PPS2_BACSU	P39846 bacillus su

34	244	7.2	563	1	4CL1_ORYSA	P17814 oryza sativ
35	238.5	7.0	683	1	ACS2_YEAST	P52910 saccharomyc
36	238.5	7.0	1274	1	SRF3_BACSU	O08787 bacillus su
37	238	7.0	537	1	4CL_PINTA	P41636 pinus taeda
38	238	7.0	672	1	ACSA_PHYBL	Q01576 phycomyces
39	236.5	7.0	548	1	LUC1_LUCCR	P13129 luciola cru
40	236	7.0	662	1	ACSA_HELPY	O25686 helicobacte
41	235.5	7.0	6486	1	TYCC_BACBR	O30409 b tyrocidin
42	234	6.9	561	1	4CL1_ARATH	Q42524 arabidopsis
43	233	6.9	652	1	ACSA_ECOLI	P27550 escherichia
44	233	6.9	683	1	ACSL_KULIA	O60011 kluyveromyc
45	231.5	6.8	660	1	ACSA_ALCEU	P31638 alcaligenes

## ALIGNMENTS

RESULT 1  
FATP\_MOUSE  
ID FATP\_MOUSE STANDARD; PRT; 646 AA.  
AC Q60714;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE LONG-CHAIN FATTY ACID TRANSPORT PROTEIN (FATP).  
GN SLC27A1 OR FATP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SWISS;  
RX MEDLINE=95042740; PubMed=7954810;  
RA Schaffer J.E., Lodish H.F.;  
RT "Expression cloning and characterization of a novel adipocyte long  
chain fatty acid transport protein.";  
RL Cell 79:427-436(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98438516; PubMed=9765271;  
RA Hui T.Y., Frohner B.I., Smith A.J., Schaffer J.E., Bernlohr D.A.;  
RT "Characterization of the murine fatty acid transport protein gene and  
its insulin response sequence.";  
RL J. Biol. Chem. 273:27420-27429(1998).  
CC -1- FUNCTION: INVOLVED IN TRANSLLOCATION OF LONG-CHAIN FATTY ACIDS  
ACROSS THE PLASMA MEMBRANE. MAY PLAY A PIVOTAL ROLE IN REGULATING  
AVAILABLE LONG-CHAIN FATTY ACID SUBSTRATES FROM EXOGENOUS SOURCES  
IN TISSUES UNDERGOING HIGH LEVELS OF BETA-OXIDATION OR  
TRIGLYCERIDE SYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PLASMA MEMBRANE.  
CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN SKELETAL MUSCLE, HEART  
AND FAT. LOWER LEVELS IN BRAIN, KIDNEY, LUNG AND LIVER. NO  
EXPRESSON IN SPLEEN OR INTESTINE.  
CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT  
COVALENT BINDING OF AMP TO THEIR SUBSTRATE.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement as long as its content is in no way  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: U15976; AAC71060.1;  
DR EMBL: AF023258; AAC69640.1;  
DR EMBL: AF023256; AAC69640.1; JOINED.  
DR EMBL: AF023257; AAC69640.1; JOINED.  
DR MGD: MGI:1347098; SLC27A1.  
DR InterPro: IPR000873;  
DR Pfam: PF0501; AMP-binding; 1.  
DR PROSITE: PS00455; AMP\_BINDING; 1.

KW Glycoprotein; Lipid transport; Transmembrane.

FT TRANSMEM 14 34 POTENTIAL.  
FT TRANSMEM 115 135 POTENTIAL.  
FT TRANSMEM 140 160 POTENTIAL.  
FT TRANSMEM 169 189 POTENTIAL.  
FT TRANSMEM 293 313 POTENTIAL.  
FT TRANSMEM 537 557 POTENTIAL.  
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 393 393 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 518 518 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 646 AA; 71276 MW; 910B92BA8D985B4C CRC64;

Query Match 62.4%; Score 2110; DB 1; Length 646;

Best local similarity 60.4%; Pred. No. 8.9e-166;

Matches 391; Conservative 97; Mismatches 152; Indels 2; Gaps 2;

QY 4 GASLVGVLLFESKL-VLKLPWTQVGSLLFLYLGSGGWRIRVFIKTRDIFGGLVLLKV 62  
DB 5 GAGTASVASLALLFWLGLPMTWSAAAFVYVGGGGFRLVIVCTARDLFGSLVLRV 64  
QY 63 KAKVROCLQERRTPILFASVTRRHDPKALIFEGTDTHTWTRQLDDEYSSSVANFLQARG 122  
DB 65 RLELRHRRAGDTIPRIFQAVARQPERALVLDASSGICWTFAQLDITYSNVANFLQGL 124  
QY 123 LASGDVAALPMENRNEFVGLWGLMAKGLVEALINTNLRDALLHCLTTSRARALVFGSE 182  
DB 125 FAPGDVAVFLEGRPEFVGLWGLAKAGVVAALLNLRREPLAFCLGTSAAKALIYGE 184  
QY 183 MASACEVHASLDPSLSLFCSSGWEPAVPPSTEHLDPKADP-KHLPSCDPKGFDTKL 241  
DB 185 MAAVAEVSQGLSKLKFCSDGLPESVLPDTQLDPLMLAEAPTTPLAQAPGKGMDDRL 244  
QY 242 FYITSGTTGLPKAAIIVVHSRYRMAALVYGYFRMPNDIVYDCLPLYHSAGNIYIGQC 301  
DB 245 FYITSGTTGLPKAAIIVVHSRYRMAALVYGYFRMPNDIVYDCLPLYHSAGNIYIGQC 304  
QY 302 LLHGMVIRKKFSASRFDDCKIYKNTIVQYIGELCRYLLNQPPEAENQHOVRMALGN 361  
DB 305 VIYGLTVVLRKKFSASRFDDCKVYKNTIVQYIGELCRYLLNQPPEAENQHOVRMALGN 364  
QY 362 KAKVROCLQERRTPILFASVTRRHDPKALIFEGTDTHTWTRQLDDEYSSSVANFLQARG 122  
DB 365 GLRPAIMEEFTQGFVQIGEFYATGECNCSIANMDKGVSGCFNSRLTHVPIRLVKV 424  
QY 422 NEDTMELIRGPDGVICPCQPGEPQLVGRILQKDPRLRRFDGYNQGANNAKAKDFKKG 481  
DB 425 NEDTMELIRGPDGVICPCQPGEPQLVGRILQKDPRLRRFDGYNQGANNAKAKDFKKG 484  
QY 482 DQAYLTGDLVMDLGYLFRDRTGTFRWKGENVSTTEVEGTLRLLDMADVAVYGV 541  
DB 485 DQAYLTGDLVMDLGYLFRDRTGTFRWKGENVSTTEVEGTLRLLDMADVAVYGV 544  
QY 542 PGTEGRAGMAVAPSTGNCNCLERFAVLEKELPLYARPIFLRLLPELHKTGYTKFQKTEL 601  
DB 545 PGVEGRAGMAAIAIDPHSQDPSNMYEQLOKVLASVAPRIFLLPQVDTGTGTFKIQKTEL 604  
QY 602 RKSGFDPAIVKDFLYLDAQKGYVPLDQEAYSRIQAGEKL 643  
DB 605 QREGFDPPQTSRFLFLDLKQGRYVPLDERVHARICAGDFSL 646

RESULT 2

FATP\_RAT

ID FATP\_RAT STANDARD; PRT; 646 AA.

AC P97849;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE LONG-CHAIN FATTY ACID TRANSPORT PROTEIN (FATP).

GN SLC27A1 OR FATP.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-98041635; PubMed-9375787;  
RA Schaap F.G., Hamers L., van der Vusse G.J., Glatz J.F.C.;  
RL "Molecular cloning of fatty acid-transport protein cDNA from rat.";  
RT Biochim. Biophys. Acta 1354:29-34(1997).  
CC -!- FUNCTION: INVOLVED IN TRANSLLOCATION OF LONG-CHAIN FATTY ACIDS  
CC ACROSS THE PLASMA MEMBRANE. MAY PLAY A PIVOTAL ROLE IN REGULATING  
CC AVAILABLE LONG-CHAIN FATTY ACID SUBSTRATES FROM EXOGENOUS SOURCES  
CC IN TISSUES UNDERGOING HIGH LEVELS OF BETA-OXIDATION OR  
CC TRIGLYCERIDE SYNTHESIS.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PLASMA MEMBRANE.  
CC -!- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT  
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U89529; AAC53424.1; -;  
DR InterPro; IPR000873; -;  
DR Pfam; PF00501; AMP-binding; 1.  
DR PROSITE; PS00455; AMP\_BINDING; 1.  
KW Glycoprotein; Lipid transport; Transmembrane.  
FT TRANSMEM 14 34 POTENTIAL.  
FT TRANSMEM 115 135 POTENTIAL.  
FT TRANSMEM 140 160 POTENTIAL.  
FT TRANSMEM 169 189 POTENTIAL.  
FT TRANSMEM 293 313 POTENTIAL.  
FT TRANSMEM 537 557 POTENTIAL.  
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 393 393 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 518 518 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 646 AA; 71283 MW; C450CF174CC2EB87 CRC64;

Query Match 61.7%; Score 2087; DB 1; Length 646;

Best Local Similarity 60.4%; Pred. No. 7.3e-166;

Matches 388; Conservative 98; Mismatches 154; Indels 2; Gaps 2;

QY 4 GASLVGVLLFESKL-VLKLPWTQVGSLLFLYLGSGGWRIRVFIKTRDIFGGLVLLKV 62  
DB 5 GAGTASVASLALLFWLGLPMTWSAAAFVYVGGGGFRLVIVCTARDLFGSLVLRV 64  
QY 63 KAKVROCLQERRTPILFASVTRRHDPKALIFEGTDTHTWTRQLDDEYSSSVANFLQARG 122  
DB 65 RLELRHRRAGDTIPRIFQAVARQPERALVLDASSGICWTFAQLDITYSNVANFLQGL 124  
QY 123 LASGDVAALPMENRNEFVGLWGLMAKGLVEALINTNLRDALLHCLTTSRARALVFGSE 182  
DB 125 FAPGDVAVFLEGRPEFVGLWGLAKAGVVAALLNLRREPLAFCLGTSAAKALIYGE 184  
QY 183 MASACEVHASLDPSLSLFCSSGWEPAVPPSTEHLDPKADP-KHLPSCDPKGFDTKL 241  
DB 185 MAAVAEVSQGLSKLKFCSDGLPESVLPDTQLDPLMLAEAPTTPLAQAPGKGMDDRL 244  
QY 242 FYITSGTTGLPKAAIIVVHSRYRMAALVYGYFRMPNDIVYDCLPLYHSAGNIYIGQC 301  
DB 245 FYITSGTTGLPKAAIIVVHSRYRMAALVYGYFRMPNDIVYDCLPLYHSAGNIYIGQC 304  
QY 302 LLHGMVIRKKFSASRFDDCKIYKNTIVQYIGELCRYLLNQPPEAENQHOVRMALGN 361  
DB 305 VIYGLTVVLRKKFSASRFDDCKVYKNTIVQYIGELCRYLLNQPPEAENQHOVRMALGN 364  
QY 362 KAKVROCLQERRTPILFASVTRRHDPKALIFEGTDTHTWTRQLDDEYSSSVANFLQARG 122  
DB 365 GLRPAIMEEFTQGFVQIGEFYATGECNCSIANMDKGVSGCFNSRLTHVPIRLVKV 424

Best Local Similarity 39.4%; Pred. No. 1e-80;  
Matches 248; Conservative 102; Mismatches 146; Indels 36; Gaps 12;

QY 422 NEDTMELIRPGDVCIPCEPQGLVGRILQKDLRRDGLYNQGNKKIAKOVFKKG 481  
DB 425 NEDTMEPLRDSQGLCIPCEPQGLVGRILQKDLRRDGLYNQGNKKIAKOVFKKG 484  
QY 482 DQAYLTGDLVMDDELGYLFRDRTGTFRWKGENVSTTEVEGTLRLLDMADVAVYGV 541  
DB 485 DSAYLSGDLVMDDELGYLFRDRTGTFRWKGENVSTTEVEGTLRLLDMADVAVYGV 544  
QY 542 PGTEGRAGNAVASPNCNCLERFAQVLEKELPYARPIFLRLPLHKTGYKFKQTEL 601  
DB 545 PGVEGSGMAAIDPHNQDLPNSMYELQKVLASQAQPIFLRLPLHKTGYKFKQTEL 604  
QY 602 RKEGFPDPAIVKDLQKGVYPLDQEAYSRIQAGEKL 643  
DB 605 QREGFPDQTSRDLFLDLKQGRYLPDLDERVHARICAGDFSL 646

RESULT 3  
VLCS\_HUMAN STANDARD; PRT; 620 AA.  
AC 014975;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE VERY-LONG-CHAIN ACYL-COA SYNTHETASE (EC 6.2.1.-) (VERY-LONG-CHAIN-  
DE FATTY-ACID-COA LIGASE).  
GN FACVLI OR VLACS.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RA Uchiyama A., Aoyama T., Kamijo K., Wakui K., Fukushima Y.,  
RA Shinozawa N., Suzuki Y., Kondo N., Orii T., Hashimoto T.;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99216327; PubMed=10198260;  
RA Steinberg S.J., Wang S.J., Kim D.G., Mihalik S.J., Watkins P.A.;  
RT "Human very-long-chain acyl-CoA synthetase: cloning, topography, and  
RT relevance to branched-chain fatty acid metabolism.";  
RL Biochem. Biophys. Res. Commun. 257:615-621(1999).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, KIDNEY, PLACENTA AND  
CC PANCREAS.  
CC -!- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT  
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; D88308; BAA23644.1; -  
DR EMBL; AF096290; AAC64973.1; -  
DR MIM; 603247; -  
DR InterPro; IPR000873; -  
DR Pfam; PF00501; AMP-binding.1.  
DR PROSITE; PS00455; AMP\_BINDING.1.  
KW Ligase; Fatty acid metabolism; Peroxisome; Transmembrane.  
FT TRANSMEM 1 21 POTENTIAL.  
FT TRANSMEM 107 127 POTENTIAL.  
FT TRANSMEM 262 282 POTENTIAL.  
SQ SEQUENCE 620 AA; 70312 MW; BFD33E1BB67B3DF8 CRC64;

Query Match

31.48; Score 1064; DB 1; Length 620;

QY 29 LFLVLGSGGWRFRVFTKTRIRDFGLVLLKVKAKVRCLOQR--RTVPILFASVRR 86  
DB 12 LFLVPL-----LVNLCPPYFQDGYFLKVAAGRRVRSYGRRPARTILRAFLERAK 65  
QY 87 HPDKTALIFEGETDTHWTFRQDDEYSSSVANFLQAR-GLASGDVAAIFWENNEEVLWL 145  
DB 66 TPHKPFLLFR--DETLTYAQVDRRSNQVARALHDLGLRQDCVALLMGNPEYVWL 123  
QY 146 MAKLGVEAALTNLRDALLHLCTTSARALVFGSEWASICEVHASL----DPSLSLFC 202  
DB 124 LVKLGCAACALNINIRAKSLHCFCCGAKVLLVSPELQAAVEILPSLKDDSYIYVS 183  
QY 203 SGWEPGAVPSTEBHLDPLLDAPKHLPSCP-----DKGFTDKLFYIVTSGTGLPKA 255  
DB 184 RTS-----NTDGLDSFL-DKVDEVSTETIPESWSEVTFSTPALYIYTSGTGLPKA 234  
QY 256 AIVVHSRYRMAALVYGFGRMRPNDIVDCLPLYHSAGNIVIGQCILLHGMVTVIRKFS 315  
DB 235 AMITHQRIWGTGLTFVS-GLKADDVYIITLPFYHSAALLIGIHGCIIVAGATLALRTKS 293  
QY 316 ASRFWDCCIKYNCITVOYIGELCRYLLNOPPREAENOHQVRLMGLNGLRQSIWTFSSRF 375  
DB 294 ASQFWDCCRYNVIVIOYIGELLYLNCSPQKPNDRDKVRLALNGLNGLDGMVVRQFVKRF 353  
QY 376 HIPOVAFYGAECNCSLGNFDSQVGCAGFNSRILSFYPIRLVVRVNEDELMIRGPDGV 435  
DB 354 GDICIEFYFAATEGNIGFMVYARKVAVGRVNYLQKKIITYDLIKYDVEKDEPVDENGY 413  
QY 436 CIPCPGPGOLVGRITIKOKPLRFDFGYLNOGA--NNKIAKDVFKKGQAYLFGDVLVMD 494  
DB 414 CVRPKGEVGLLVCKITQLTP---FNGYAKAQAQTEKKLRDVKFGDGLYFNSGDLMLVD 470  
QY 495 ELGYLFRDRTGDTFRWKGENVSTTEVEGTLRLLDMADVAVYGVPEGTEGRAGMAVA 554  
DB 471 HENFIYFHDVGRDTRFRWKGENVSTTEVEGTLRLLDMADVAVYGVPEGTEGRAGMAVA 530  
QY 555 SPTGN-COLERFAQVLEKELPYARPIFLRLPLHKTGYKFKQTELKKEGDFPAIVKD 613  
DB 531 MKENHEFDGKLLFOHIADYLPYARPRFLRITQDTIETGTGFKHKKMTLVEEGFNPAVKD 590  
QY 614 PLFYLDQAQGRYVPLDQEAYSRIQAGEKL 643  
DB 591 ALIFLDDTAKMVPWTEIDIYNAISAKTLKL 620

RESULT 4  
VLCS\_RAT STANDARD; PRT; 620 AA.  
AC P97524;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE VERY-LONG-CHAIN ACYL-COA SYNTHETASE (EC 6.2.1.-) (VERY-LONG-CHAIN-  
DE FATTY-ACID-COA LIGASE).  
GN FACVLI OR VLACS.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=WISTAR; TISSUE=Liver;  
RX MEDLINE=97094763; PubMed=8939997;  
RA Uchiyama A., Aoyama T., Kamijo K., Uchida Y., Kondo N., Orii T.,  
RA Hashimoto T.;  
RT "Molecular cloning of cDNA encoding rat very long-chain acyl-CoA  
RT synthetase.";  
RL J. Biol. Chem. 271:30360-30365(1996).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.  
CC -!- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT

CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; D85100; BAA12722.1; -  
 DR InterPro; IPR000873; -  
 DR Pfam; PF00501; AMP-binding; 1.  
 DR PROSITE; PS00455; AMP-BINDING; 1.  
 KW Ligase; Fatty acid metabolism; Peroxisome; Transmembrane.  
 FT TRANSMEM 1 21 POTENTIAL.  
 FT TRANSMEM 107 127 POTENTIAL.  
 FT TRANSMEM 262 282 POTENTIAL.  
 SQ SEQUENCE 620 AA; 70693 MW; 6CF9362DC3805526 CRC64;

Query Match 30.9%; Score 1044; DB 1; Length 620;  
 Best Local Similarity 38.0%; Pred. No. 4.8e-79;  
 Matches 237; Conservative 113; Mismatches 241; Indels 32; Gaps 13;

QY 32 LYLGGSGWRFIRVIKT-----IRRDFGGLVLLKAKVQCLOER--RTVPILFASTVR 85  
 DB 5 LYTGLAGLLLLPCCPYLLQDVRFLLQANMARQVRSYRQRRPVRTILHVFLEQAR 64  
 QY 86 RHPDKTALIFEGETDTHWTFQLDDEYSSVANFLQAR-GLASGDVAAIFMENNEFVGLWL 144  
 DB 65 KTHPKPFLFR--DETLTYAQVDRRSNOVARALHDHGLRGQDCVAFMGNEPAYVWLWL 122  
 QY 145 GMAKLGVEAALINTLRDALLHCLTTSRARALVFGSEMASAICEVHASL-DPSLSLFC 202  
 DB 123 GLLKLGCPMACLNTNIRAKSLHCFQCCGAKVLLASPELHEAVEEVLPLKKEGVSVFV 182  
 QY 203 SGSEWEGAVPPSTEHLPLK--APKHLSPCDKGTDFKLFYIYTSCTTGLPKAAIVVHS 261  
 DB 183 SRTSNNGVDTILDKVDGVSADIPESWRS--EVTFTTTPAVYIYTSCTTGLPKAAITNH 240  
 QY 262 RYRMAALVYGFPM-----RPNDIVYDCLPLYHSAGNIVGIGQCLLHGMVWIRKFS 315  
 DB 241 R-----LWIGSLARSGIKAHDIYITMPLYHSAALMIGLHGCIVVGATFALRSKFS 293  
 QY 316 ASRFWDCKIKYNTVIQYIGELCRYLLNQPPREAHQHVQRMALGNLGRQSIWTFSSRF 375  
 DB 294 ASQFWDCKRYNATVIQYIGELRLYLNTPOKPNDRDHVKVIALGNLGRGVWREFIKRF 353  
 QY 376 HIPOVAEFGATECNSLGNFDSQVACGNSKILSFVYPIRLVRVNETMELIRGPDGV 435  
 DB 354 GDHIVEFYASTEGNIGFMNYPKIGAVGRENLYQKKVVRHKLKVDKDEPVRDANGY 413  
 QY 436 CIPQCPGEPQVLGRITQKPLRRPDGYL--NQANNKKIAKDVFKKQDAVLTGDVLMVD 494  
 DB 414 CIKPRGEVGLLCKITELTP---FPGYAGGKTQTEKKLRDVKFGDVFNSGDLMLD 470  
 QY 495 ELGYLPRDRGTDFRWKGNSTTEVEGTLRLLDMADVGVVGPTEGREGAGMAVA 554  
 DB 471 RENFIYFDRVGTDFRWKGNVATVADIVGLVDFVEEVENYGVDPVPGHEGRIGMASIK 530  
 QY 555 SPTG-NCDLERFAQVLEKELPLYARITFLRLPELHKTCTYKFKTELKKEGDFDAIVKD 613  
 DB 531 MKENYFNGKKLFQHSIEYLPSRSRFRURIQDTIEITGTFRHKRVTLMEEGFNPSVIKD 590  
 QY 614 PLFYDLAQKGRYVPLDQEAYSRI 636  
 DB 591 TLFMDDTTEKTYVPMTEIDIYAI 613

RESULT . 5  
 VLCS\_MOUSE  
 ID VLCS\_MOUSE STANDARD; PRT; 620 AA.

AC O35488;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE VERY-LONG-CHAIN ACYL-COA SYNTHETASE (EC 6.2.1.-) (VERY-LONG-CHAIN-  
 DE FATTY-ACID-COA LIGASE).  
 GN FACVL1 OR VLACS OR VLCS.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TissUP=Liver;  
 RA Kemp S., Lu J.-F., Smith K.D.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.  
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT  
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF033031; AAB87982.1; -  
 DR InterPro; IPR000873; -  
 DR Pfam; PF00501; AMP-binding; 1.  
 DR PROSITE; PS00455; AMP-BINDING; 1.  
 KW Ligase; Fatty acid metabolism; Peroxisome; Transmembrane.  
 FT TRANSMEM 1 21 POTENTIAL.  
 FT TRANSMEM 107 127 POTENTIAL.  
 FT TRANSMEM 262 282 POTENTIAL.  
 SQ SEQUENCE 620 AA; 70366 MW; 77C98BD0DE3B9FFB CRC64;

Query Match 30.2%; Score 1022; DB 1; Length 620;  
 Best Local Similarity 37.7%; Pred. No. 3.2e-77;  
 Matches 235; Conservative 108; Mismatches 248; Indels 32; Gaps 12;

QY 32 LYLGGSGWRFIRVIKT-----IRRDFGGLVLLKAKVQCLOER--RTVPILFASTVR 85  
 DB 5 LYTGLAGLLLLPCCPYLLQDVRFLLQANMARQVRSYRQRRPVRTILHVFLEQAR 64  
 QY 86 RHPDKTALIFEGETDTHWTFQLDDEYSSVANFLQAR-GLASGDVAAIFMENNEFVGLWL 144  
 DB 65 KTHPKPFLFR--DETLTYAQVDRRSNOVARALHDHGLRGQDCVAFMGNEPAYVWLWL 122  
 QY 145 GMAKLGVEAALINTLRDALLHCLTTSRARALVFGSEMASAICEVHASL--DPSLSLFC 202  
 DB 123 GLLKLGCPMACLNTNIRAKSLHCFQCCGAKVLLASPELHEAVEEALPTLKDAVSFVY 182  
 QY 203 SGSEWEGAVPPSTEHLPLK--LKDAPKHLSPCDKGTDFKLFYIYTSCTTGLPKAAIVVHS 261  
 DB 183 SRTSNNGVDTILDKVDGVSADIPESWRS--EVTFTTTPAVYIYTSCTTGLPKAAITNH 240  
 QY 262 RYRMAALVYGFPM-----RPNDIVYDCLPLYHSAGNIVGIGQCLLHGMVWIRKFS 315  
 DB 241 R-----LWYGTGLAMSSGITAQDVITMPLYHSAALMIGLHGCIVVGATLALRSKFS 293  
 QY 316 ASRFWDCKIKYNTVIQYIGELCRYLLNQPPREAHQHVQRMALGNLGRQSIWTFSSRF 375  
 DB 294 ASQFWDCKRYNATVIQYIGELRLYLNTPOKPNDRDHVKVIALGNLGRGVWREFIKRF 353  
 QY 376 HIPOVAEFGATECNSLGNFDSQVACGNSKILSFVYPIRLVRVNETMELIRGPDGV 435  
 DB 354 GDHIVEFYASTEGNIGFMNYPKIGAVGRENLYQKKVVRHKLKVDKDEPVRDANGY 413  
 QY 436 CIPQCPGEPQVLGRITQKPLRRPDGYL--NQANNKKIAKDVFKKQDAVLTGDVLMVD 494  
 DB 414 CIKPRGEVGLLCKITELTP---FPGYAGGKTQTEKKLRDVKFGDVFNSGDLMLD 470  
 QY 495 ELGYLPRDRGTDFRWKGNSTTEVEGTLRLLDMADVGVVGPTEGREGAGMAVA 554  
 DB 471 RENFIYFDRVGTDFRWKGNVATVADIVGLVDFVEEVENYGVDPVPGHEGRIGMASIK 530  
 QY 555 SPTG-NCDLERFAQVLEKELPLYARITFLRLPELHKTCTYKFKTELKKEGDFDAIVKD 613  
 DB 531 MKENYFNGKKLFQHSIEYLPSRSRFRURIQDTIEITGTFRHKRVTLMEEGFNPSVIKD 590  
 QY 614 PLFYDLAQKGRYVPLDQEAYSRI 636  
 DB 591 TLFMDDTTEKTYVPMTEIDIYAI 613

Db 414 C1KVPKGEVLLVCKITQLTP---FIGYAGGKTQTEKKKLKRDVFKKGGDIYFNSGDLMLD 470  
 QY 495 ELGVLXPRDSTGDTFRKKGENVSTTEVEGTLRLDLMDADVGVVPGTEGRAGMAAV- 553  
 Db 471 RENFVYFHRDVGDTFRKKGENVSTTEVEGTLRLDLMDADVGVVPGTEGRAGMAAV- 530  
 QY 554 ASPTGNCGLDERFAVLEKELEPLFYARFLLRLPELHKTGYFVKFTELKRGDFDAIVKD 613  
 Db 531 IKENYFNGKKLQHIAEYLPYARFLLRLPELHKTGYFVKFTELKRGDFDAIVKD 590  
 QY 614 PLFYLDKAGRYVPLDQAYSRI 636  
 Db 591 TLYFMDDAEKTFFVPMENIYNAI 613

## RESULT 6

FAT1\_YEAST  
 ID FAT1\_YEAST STANDARD; PRT 623 AA.  
 AC P38225;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE PROBABLE LONG-CHAIN FATTY ACID TRANSPORT PROTEIN.  
 GN FAT1 OR YBR041W OR YBR041L.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 RN NCBI\_TaxID=4932;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RA Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A.,  
 RA Viissers S.;  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP CHARACTERIZATION.  
 RC STRAIN=W303A;  
 RX MEDLINE=97236810; PubMed=9079682;  
 RA Faergeman N.J., Dirusso C.C., Elberger A., Knudsen J., Black P.N.;  
 RT "Disruption of the Saccharomyces cerevisiae homologue to the murine  
 RT fatty acid transport protein impairs uptake and growth on long-chain  
 RT fatty acids";  
 RL J. Biol. Chem. 272:8531-8538(1997).

CC -!- FUNCTION: MAY BE INVOLVED IN LONG-CHAIN FATTY ACIDS UPTAKE, AND  
 CC PRIOR TO METABOLIC UTILIZATION. MAY PLAY AN IMPORTANT ROLE IN  
 CC UPTAKE OF THESE HYDROPHOBIC COMPOUNDS UNDER CONDITIONS WHERE FATTY  
 CC ACID SYNTHESIS IS COMPROMISED.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -!- SIMILARITY: TO ENZYMES WHICH ACT VIA AN ATP-DEPENDENT COVALENT  
 CC BINDING OF AMP TO THEIR SUBSTRATE.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: Z35910; CAAB4983.1; -  
 CC DR PIR: S45899; S45899.  
 CC DR SGD: S0000245; FAT1.  
 CC DR InterPro: IPR000873; -  
 CC DR Pfam: PF00501; AMP-binding; 1.  
 CC DR PROSITE: PS00455; AMP\_BINDING; 1.  
 CC KW Lipid transport; Transmembrane.  
 CC FT TRANSMEM 6 26 POTENTIAL.  
 CC FT TRANSMEM 54 71 POTENTIAL.  
 CC FT TRANSMEM 149 169 POTENTIAL.  
 CC FT TRANSMEM 293 313 POTENTIAL.  
 CC FT CARBOHYD 184 184 N-LINKED (GLCNAC... ) (POTENTIAL).  
 CC FT CARBOHYD 289 289 N-LINKED (GLCNAC... ) (POTENTIAL).  
 CC FT CARBOHYD

FT CARBOHYD 534 534 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 591 591 N-LINKED (GLCNAC... ) (POTENTIAL).  
 SQ SEQUENCE 623 AA; 71697 MW; 0AE02700B60C8CFE CRC64;

Query Match 21.1%; Score 714; DB 1; Length 623;  
 Best Local Similarity 31.8%; Pred. No. 1.4e-51;  
 Matches 205; Conservative 92; Mismatches 260; Indels 88; Gaps 19;

QY 1 MLLGASLVGLLFSKLVKLPWTOVGSLLFL-----YLGSGGWRFFIRFKT- 48  
 Db 7 VVFLSRIFLLLR--LIKLIITPIQSLGLYFGNYDFELDKRYKEDWTIIPYFLKS 64  
 QY 49 -----IRRDIFGGLVLLKVRKQCLQERRTPILFASTVRRHRPDKTALIF-----E 96  
 Db 65 FCYIIDVRRHRFQNW-----YLFKQVQONGDHLAISYRPMK 104  
 QY 97 GTDTHWTFRLDEYS-----SSVANFLQARGLASGDVAIAFMENRNFVGLWMAKLV 151  
 Db 105 GFOLETFYIETYNIVLRSLHILHF--DYNVQAGDYVAIDCTNKLFLVFLWLSLWIGA 162  
 QY 152 EAALINTNLRDALLHCLITTSRARALVFGSEMAAICEVHASLDPSLSLFCSGSWEPGAV 211  
 Db 163 IPAFINYNKGTPLVHSLKISNITQVDFIDPASNPIRESEEEKNAL-----PDVK 213  
 QY 212 PPSTEHLID-----PLLKADPKHLPSCPDK---GFTD--KLFYIYTSGTGLPKAAIVVHSR 262  
 Db 214 LNYLEEQDLMEHLLNSQSPEFFLOQDNVRLPLGLTDFKPSMLIVTSGTGLPKSAINSWRK 273  
 QY 263 YRMAALVYVYGRMRPNDIVDCLPLVHSAGNIVGICQLLHGMVTVIRKFFSASRWDD 322  
 Db 274 SSVCCQVFGVHLVHNSTVFTAMPFHSHTAALLGACAILSHGGCLALSFKFSASTFWKQ 333  
 QY 323 CIKYNCTIVQYIGELCRYLLNQPPREAEHQVRLMALNGLSIWTNFSRSHIPOVAE 382  
 Db 334 VYLTGATHIQYVGEVCRVYLLHTPISEKMKVYKVAINGRLRDPDIQDFRKNIEVIGE 393  
 QY 383 FYGATECNSLGNF---DSQVGCAGFNSRILSFVYPIR--LVRV--NEDPMELIRPGDV 435  
 Db 394 FYAETAPFATTTTQKGDGFGIGACRNYGTIIQWFLSQOTLVRRMDPNDSDV-IYRNSKGF 452  
 QY 436 CIPQCEPGEOLVGRI-IQKDLRRRDPGYLNOGANNK-KIAKDVFKKGDQAYLTGDVLYM 493  
 Db 453 CEVAPVGEPEMLRIFFPKPKPETSQYGLGNKAKETSKVVRDVRFRGDWYRCGDLILA 512  
 QY 494 DELGILYFRDRTGDTFRWKGENVSTTEVEG--TSLRLDMADVAVYGVVPGTEGRAGMA 551  
 Db 513 DEYGLWYFLDRMGDTFRWKSENVSTTEVEQDLTASNKEQYQVLYVVGKVPKYEGRAGFA 572  
 QY 552 AVASPTGNCGLERFAQVLEK-----ELPLYARPIFLRLPELHKT 591  
 Db 573 VIKLTDNSLDITAKTLKLLNDSLSRLNLPYAMPFLFVKFVDEIKMT 617

## RESULT 7

CAIC\_ECOLI  
 ID CAIC\_ECOLI STANDARD; PRT; 522 AA.  
 AC P31552;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE PROBABLE CROTONOBETAINE/CARNITINE-COA LIGASE (EC 6.3.2.-).  
 GN CAIC.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OC NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC STRAIN=044\_K74;  
 RX MEDLINE=95115348; PubMed=7815937;  
 RA Eichler K., Bourgis F., Buchet A., Kleber H.-P.,  
 RA Mandrand-Berthelot M.-A.;









GN 4CL2 OR 4CL-2.  
 OS Solanum tuberosum (Potato).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
 OC Solanales; Solanaceae; Solanum.  
 OX NCBI\_TaxID=4113;  
 RN [1]  
 RP MEDLINE-91217100; PubMed-2022667;  
 RX Becker-Andre M., Schulze-Lefert P., Hahlbrock K.;  
 RT "Structural comparison, modes of expression, and putative cis-acting  
 RT elements of the two 4-coumarate: CoA ligase genes in potato.";  
 RL J. Biol. Chem. 266:8551-8559(1991).  
 CC -1- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE +  
 CC 4-COUMAROYL-COA.  
 CC -1- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID  
 CC METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.  
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT  
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.  
 DR PIR; B39827; B39827.  
 DR HSP; P08659; ILCI.  
 DR InterPro: IPR000873;  
 DR Pfam: PF00501; AMP-binding; 1.  
 DR PROSITE: PS00455; AMP BINDING; 1.  
 KW Ligase; Phenylpropanoid metabolism; Multigene family.  
 SQ SEQUENCE 545 AA; 59625 MW; 5481F0B0AFAEA39E0 CRC64;

Query Match 8.9%; Score 300; DB 1; Length 545;  
 Best Local Similarity 24.1%; Pred. No. 3.5e-17;  
 Matches 132; Conservative 96; Mismatches 230; Indels 90; Gaps 21;

QY 90 KTALEFGTDTHTFQRLDEYSSVANFLQARGLASGDVAALPMENRNEFVGLTGMKAL 149  
 DB 44 RCLDGDANDRIYTYAEVELTKRKVAVGLNKLGIQKQDTIMILLPNCPEFVFAFGASYL 103  
 QY 150 GVEAALINTNLRDALLHCLTTSRARALVFGSEMAAICEVHASLDPSLSFCSSGWEPG 209  
 DB 104 GAISTMANPLFTAEEVVKQAKASSAKIVITQACFAGKVDY--AIENDLKVICVDSAPG 161  
 QY 210 ANP-----PSTEHLDPKADPHLPCSPDKGFTDKLFYITSGTGLPKAAIVVHSRY 264  
 DB 162 CVHSELIQSDHEIPDVKIQPDVVALP-----YSSGTTGLPKGVMTHKGLV 210  
 QY 265 RMAALVYVG---FRMRPNDIVDCLPLVHSAGNIVIGQCILL-----HGMTVVIKKFSA 316  
 DB 211 TSAQQVDGENANLYMHSDDVLMCVLPFH-----IYSLNSVLLCALRYGAAILIMQKFDI 266  
 QY 317 SRFWDDCIKYNCTIVQYIGELCRYLLNQPPEAENQHVMA-----LGNGLRQSIWN 370  
 DB 267 AQFLELIPKHKVTIGFVPPVIVLATAKSPLVHNYDLSSRVTVMSGAAPLGLEDAV--- 323  
 QY 371 FSSRPHIQVAEYFGATECNSLG-----NFDQVGCAGFNSRILSFYVPIRLVRN 422  
 DB 324 -RAKFPNKLGGYGTGTEAGPVLMCLAFAPKPEFDIKSGAGC-----TVVRNA 370  
 QY 423 EDTMELIRGPGVCITPCQPGEGVLVRIQKDLRRFDGVLNQGANNKKTAKDVFKGD 482  
 DB 371 E--MKIVDPDTCGLP--RNQPGEI---CIRGDQIMK--GYLN-----DPEATARTIEKG- 417  
 QY 483 QAYL-TGVDVLMDELGYLFFDRDGTDFRWKGENVSTTEVEGTLRLDMDADVAYGEV 541  
 DB 418 --WLHTGIDGIDDDDELFDVDRUKELLYKGFQVAPAEALLINHPDIDDAV---V 471  
 QY 542 PTEGRAGMAVA-----SPTGNCDLERFAQVLEKELPLY---ARPIFLRLPELHKGTGY 594  
 DB 472 PMIDEQAGEVPVAFVVRNSNGSTITEDEVKDFISKQVIFYKRIKRVFFETVP---KSPSG 528  
 QY 595 KFOKTEL 602  
 DB 529 KILRKDL 536

RESULT 12  
 4CL1 TOBAC  
 ID 4CL1 TOBAC STANDARD; PRT; 547 AA.  
 AC 024145;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE 4-COUMARATE--COA LIGASE 1 (EC 6.2.1.12) (4CL 1) (4-COUMAROYL-COA  
 DE SYNTHASE 1).  
 DE 4CL1.  
 GN Nicotiana tabacum (Common tobacco).  
 OS Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
 OC Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-96416441; PubMed-8819324;  
 RA Lee D., Douglas C.J.;  
 RT "Two divergent members of a tobacco 4-coumarate:coenzyme A ligase  
 RT (4CL) gene family. cDNA structure, gene inheritance and expression,  
 RT and properties of recombinant proteins.";  
 RL Plant Physiol. 112:193-205(1996).  
 CC -1- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE +  
 CC 4-COUMAROYL-COA.  
 CC -1- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID  
 CC METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.  
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT  
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U50845; AB18637.1; -  
 DR HSP; P08659; ILCI.  
 DR Mandel; 24477; Nicta; 1179; 24477.  
 DR InterPro: IPR000873;  
 DR Pfam: PF00501; AMP-binding; 1.  
 DR PROSITE: PS00455; AMP BINDING; 1.  
 KW Ligase; Phenylpropanoid metabolism; Multigene family.  
 SQ SEQUENCE 547 AA; 59842 MW; 154DF6D684E3F51B CRC64;

Query Match 8.8%; Score 299; DB 1; Length 547;  
 Best Local Similarity 24.0%; Pred. No. 4.2e-17;  
 Matches 132; Conservative 93; Mismatches 232; Indels 92; Gaps 22;

QY 90 KTALEFGTDTHTFQRLDEYSSVANFLQARGLASGDVAALPMENRNEFVGLTGMKAL 149  
 DB 46 RCLINGANDRIYTYAEVELTKRKVAVGLNKLGIQKQDTIMILLPNSPEFVFAFGASYL 105  
 QY 150 GVEAALINTNLRDALLHCLTTSRARALVFGSEMAAICEVHASLDPSLSFCSSGWEPG 209  
 DB 106 GAISTMANPLFTAEEVVKQAKASSAKIIITQSCFVKVDYASEND--VKVICDSAPG 163  
 QY 210 A-----VPPSTEHLDPKADPHLPCSPDKGFTDKLFYITSGTGLPKAAIVVHSRY 264  
 DB 164 CLHFSELTSQDEHEIPEVKIQPDVVALP-----YSSGTTGLPKGVMTHKGLV 212  
 QY 265 RMAALVYVG---FRMRPNDIVDCLPLVHSAGNIVIGQCILLHM-----TVVIRKFS 316  
 DB 213 TSAQQVDGENANLYMHSDDVLMCVLPFH-----IYSLNSVLLCALRYGAAILIMQKFDI 268  
 QY 317 SRFWDDCIKYNCTIVQYIGELCRYLLNQPPEAENQHVMA-----LGNGLRQSIWN 370  
 DB 269 APFLELIQKVKYSGFVPPVIVLATAKSPDIVDSSVTVMSGAAPLGLEDAVRTK 328  
 QY 371 FSSRPHIQVAEYFGATECNSLG-----NFDQVGCAGFNSRILSFYVPIRLVRN 422

Db 329 FPN-----AKLGCGYGTGAGPVLAMCLAFAPKPFIDKSGACG-----TWRNA 372  
 QY 423 EDTMELIRGPDGVCIPQCGPGLVRIQKPLRRFDGYLNO--GANNKTIKADVFKG 481  
 Db 373 E--MKIVDPDTGSLP--RNQFGEI---CIRGDIKMK--GYLNDPEATRTDK-----417  
 QY 482 DQAYL--TGDVLMDELGYLYFRDRTGDTFRWKGENVSTTEVEGTLRLDMDADVAVGYE 540  
 Db 418 -EGWLTGDIIGFIDEDELFIYDRLKELIKYGFQVAPAEIEALLNHPNISDAV---472  
 QY 541 VPGTEGRAG---MAAVASPTGNCLE-REFAQVLEKELPLY---ARPIFLRLPELHKGT 593  
 Db 473 VPMKDEQAGEVPAFVVRNSGSAITDEVKDFISKQVIFKRVKRVFFVETVP---KSPS 529  
 QY 594 YKFOKTELR 602  
 Db 530 GKILRKDLR 538

RESULT 13  
 4CL3\_ARATH STANDARD; PRT; 561 AA.  
 AC 995777;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE 4-COUMARATE--COA LIGASE 3 (EC 6.2.1.12) (4CL 3) (4-COUMAROYL--COA  
 DE SYNTHASE 3)  
 GN 4CL3 OR ATIG65060 OR F16G16.6.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=99348176; PubMed=10417722;  
 RA Ehling J., Buettner D., Wang Q., Douglas C.J., Somssich I.E.,  
 RA Kombrink E.;  
 RT "Three 4-coumarate:coenzyme A ligases in Arabidopsis thaliana  
 RL represent two evolutionarily divergent classes in angiosperms.";  
 RL Plant J. 19:9-20(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Mafti R., Marziali A.,  
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzer S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 thaliana.";  
 RL Nature 408:816-820(2000).  
 CC -1- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE +  
 CC 4-COUMAROYL--COA.  
 CC -1- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOIC  
 CC METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.  
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT  
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to a license@sib-sib.ch).  
 CC -----  
 DR EMBL; AF106087; AAD47194.1; -  
 DR EMBL; AF106088; AAD47195.1; -  
 DR EMBL; AC009360; AAF06039.1; -  
 DR HSP; P08659; ILCI.  
 DR InterPro; IPR000873; -  
 DR Pfam; PF00501; AMP-binding; 1.  
 DR PROSITE; PS00455; AMP\_BINDING; 1.  
 KW Ligase; Phenylpropanoid metabolism; Multigene family.  
 SQ SEQUENCE 561 AA; 61310 MW; C2EFF1C36B33F6DC CRC64;

Query Match 8.7%; Score 293.5; DB 1; Length 561;  
 Best Local Similarity 25.3%; Pred. No. 1.3e-16;  
 Matches 139; Conservative 84; Mismatches 237; Indels 89; Gaps 23;  
 QY 89 DKTALIFEGDTHWTFQRLDEYSSVANFLQAGLSADYAAIFMENNEFVGLWLGMAK 148  
 Db 63 DKPCLIVGSGTSYTGETHLICRRVASGLYKLGIRKGDVIMILLQNSAEVFSFGASM 122  
 QY 149 LGVEAALINTNLRRDALLHCLTTSRARALVFGSEMAICEVHASLPSLSLFCSGSWEP 208  
 Db 123 IGAVSTANPYTSQELYKOLKSSGAKLIITHSQYVDKL---KNLGNLTITDDETP 178  
 QY 209 -GAVPPSTEHLDPKLPKPCPKDGKFTDKLYFTYTGTLGPKAAIIVHSRYRMA 267  
 Db 179 ENCLPFTSLITD---DETNPFTQVTDIGDDAAALPFSSTGTLGPKGVVLTHKSLITSV 234  
 QY 268 ALVYVG---FRMRPNDIVDCLPLYSAGNIVGICQLLH---GMTVVRKKFSASRF 319  
 Db 235 AQQVDGDNPNLYKSNVDILCVLPFH---IYSLNSVLLNSLRSGLATVLLMHKFEIGAL 290  
 QY 320 WDDCIKYNCTIVQYIGELCRYLLNPPREANQHOVRMA-----LGNLRQSIWTNFS 373  
 Db 291 LDLIQRHRTVIAALVPLVIALAKNPTVNSYDLSSVFLSGAAPLKGELQDSL-----344  
 QY 374 RHPHQ--VAEYFGATEC---NCSLG---NFDQVACGACFNSRILSFVYPIRLVRVNE 423  
 Db 345 RRRPQAILGQYGTGTEAGPVLMSLGFAPKPIPTKSGSCG-----TVVRNAELKVVHL 398  
 QY 424 DTMELIRG---PDGVCIPQCPGQVLVRIQKPLRRFDGYLNO--GANNKTIKADVFK 479  
 Db 399 ET-RLSLGYNQPGECIR-----GQIMKE-----YLNDEPATSATI-----434  
 QY 480 KGDQAYL--TGDVLMDELGYLYFRDRTGDTFRWKGENVSTTEVEGTLRLDMDADVAVYG 538  
 Db 435 -DEEGWLTGDIIGYVDEDEIFIVDRKEVYKFKGFQVPPAEELSLNHHHSIAADA--491  
 QY 539 VEVPGTEGRAG---MAAVASPTGNCLE-REFAQVLEKELPLYARPIFLRLPELHKGT 593  
 Db 492 --VPONDEVAGEVPAFVVRNSGN-DITEEDVKEVAKQVYFKLHKVFFVASIPKSPS 548  
 QY 594 YKFOKTELR 602  
 Db 549 GKILRKDLK 557

RESULT 14  
 4CL3\_PETCR STANDARD; PRT; 544 AA.  
 AC P14912;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE 4-COUMARATE--COA LIGASE 1 (EC 6.2.1.12) (4CL 1) (4-COUMAROYL--COA



DR Pfam; PF00501; AMP-binding; 1.  
DR PROSITE; PS00455; AMP\_BINDING; 1.  
DR Ligase; Phenylpropanoid metabolism  
KW SEQUENCE 544 AA; 59783 MW; B47

Query Match 8.3%; Score 282; DB 1; Length 544;

[illegible]

Search completed: July 16, 2001, 18:15:13  
Job time: 189 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 18:15:11 ; Search time 18.37 seconds  
(without alignments)  
1199.035 Million cell updates/sec

Title: US-09-405-504A-53  
 Perfect score: 3384  
 Sequence: 1 MLLGASLVGVLFESKLVKL.....RYVPLDOEAYSRTOAGEEKL 643

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues  
Total number of hits satisfying chosen parameters: 93435

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			ID	Description
	Score	Match	Length DB		
1	2110	62.4	646	1	FATP_MOUSE Q60714 mus musculus
2	2087	61.7	646	1	FATP_RAT P97849 rattus norv
3	1064	31.4	620	1	VLC5_HUMAN O14975 homo sapien
4	1044	30.9	620	1	VLC5_RAT P97524 rattus norv
5	1022	30.2	620	1	VLC5_MOUSE O35488 mus musculus
6	714	21.1	623	1	FAT1_YEAST P38225 saccharomyc
7	350.5	10.4	522	1	CATC_ECOLI P1552 escherichia
8	318	9.4	661	1	ACSA_COPCI O13440 coprinus ci
9	311.5	9.2	561	1	LCPA_ECOLI P29212 escherichia
10	303	9.0	545	1	CLL1_SOLTU P31684 solanum tub
11	300	8.9	545	1	CLL2_SOLTU P31685 solanum tub
12	299	8.8	547	1	CLL1_TOBAC O24145 nicotiana t
13	293.5	8.7	541	1	CLL3_ARATH P95777 arabidopsi
14	286	8.5	544	1	CLL1_PETRC P14912 petroselinu
15	282	8.3	544	1	CLL2_PETCR P14913 petroselinu
16	278	8.2	553	1	CLL_VANPL O24540 vanilla pla
17	271.5	8.0	562	1	LCPA_HAEIN P46450 haemophilus
18	271	8.0	543	1	FAT2_YEAST P38137 saccharomyc
19	270.5	8.0	656	1	ACSA_SHOCA O68040 rhodobacter
20	267.5	7.9	542	1	CLL2_TOBAC O24146 nicotiana t
21	262	7.7	694	1	ACSA_CRYPV Q27549 cryptospori
22	260	7.7	5255	1	BACA_BACLI P68006 b bacitraci
23	259	7.7	670	1	ACSA_EMENI P15928 emeritella
24	259	7.7	3587	1	SRLF_BACSU P27206 bacillus su
25	255.5	7.6	3587	1	SRLF_BACSU P27206 bacillus su
26	255	7.5	669	1	ACSA_PENCH Q04747 bacillus su
27	253	7.5	6359	1	BACC_BACLI P36333 penicillium
28	251	7.4	556	1	CLL2_ARATH O68008 b bacitraci
29	250	7.4	3587	1	TYCB_BACBR Q95725 arabidopsi
30	249	7.4	548	1	YDID_ECOLI O30408 b tyrocidin
31	247	7.3	675	1	ACSA_CANAL P38135 escherichia
32	246	7.3	713	1	ACS1_YEAST O94049 candida alb
33	246	7.3	2560	1	PPS2_BACSU Q01574 saccharomyc
34	246	7.3	2560	1	PPS2_BACSU P39846 bacillus su

FT	TRANSMEM	14	34	POTENTIAL.
FT	TRANSMEM	115	135	POTENTIAL.
FT	TRANSMEM	140	160	POTENTIAL.
FT	TRANSMEM	169	189	POTENTIAL.
FT	TRANSMEM	293	313	POTENTIAL.
FT	TRANSMEM	537	557	POTENTIAL.
FT	CARBOHYD	330	330	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	393	393	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	518	518	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	646 AA;	71276 MM;	910B92BA8D985B4C CRC64:

[illegible]

RAID	FATP_RAT	STANDARD;	PRT;	646 AA.
AC	P97849;			
TC	01-NOV-1997	(Rel. 35, Created)		
TT	01-NOV-1997	(Rel. 35, Last sequence update)		
TT	30-MAY-2000	(Rel. 39, Last annotation update)		
EE	LONG-CHAIN FATTY ACID TRANSPORT PROTEIN (FATP).			
EE	SLC27A1 OR FATP.			
OS	Rattus norvegicus (Rat).			
OS	Fukushima; Matsuzaki; Chiba; et al.			



Best Local Similarity 39.4%; Pred. No. 1e-80;  
Matches 248; Conservative 102; Mismatches 244; Indels 36; Gaps 12;

```

QY 29 LFLYLGSGGWRIRVFIKTRIRDFGLVLLKAKVRQCLQER--RTVPILFASTVRR 86
D 12 LFLPL-----LVNCCPYFFODIGYFLKVAAGVRVRSYQRRPARTILRAFLKARQ 65
QY 87 HPDKTALIFEGTHTHFERQDEYSSSVANFLOAR--GLASGDVAALFMRNNEFFVGLWIG 145
D 66 TPHPKPELLFR--DETUTYAQVDRRSNQVARALHDHGLRQGDVALLMGNEPAYVWLWIG 123
QY 146 MAKLGVEAALINTNRDRLHCLTTSRARALVFGESEMAAICEVHASL---DPSLSLFC 202
D 124 LVKLGCAACLVNIRAKSLLHCFCCGAKVLLVSPQLQAQVEELPSLKDDVYIYVS 183
QY 203 SGSEWEGAVPSTHELDPLKAPKHPSCP-----DKGTDLKFIYITSGTGLPKA 255
D 184 RTS-----NTDGDLSFL-DKVDVSTEPPIPEISWSEVTFSTPALYITSTGTLPLKA 234
QY 256 AIVVHSRYRMAALVYGFMRPNIDIVDCLPLYHSAGNIVGIGOCLLHGMTVIRKKFS 315
D 235 AMITHQRIWGTGLTFVS--GLKADVDIYITLFPYHSAALLIGICIVAGATLARTKFS 293
QY 316 ASRFWDCCIKNCTIVQYIGELCRYLLNQPPREAEHQVRLMALGNLGRSQSTWTFSSRF 375
D 294 ASQFWDCCRKNYNTVQYIGELRLYLCSNPKPDRDHKVRALGNLGRDWRQFVKRF 353
QY 376 HIPOVAEFYATGECNCSLGNEDSOVACGFSNRLSFIYVPIRLVRYNEDTMELIRGPDV 435
D 354 GDICIYFAATGEGNIFGNFYARKYAGVRVNYLQKKIITLYLDIYDEKDEPVRDENG 413
QY 436 CIPQCPGEGVLGRRIQKDLRRFDGYLNOGA--NNKIAKDVKFKGQDQAVLTGDLVMD 494
D 414 CVRPKGEVGLLVCKITQLTLP---ENGAYAKAQTEKKLRDVKFGDLYNSGDLWMD 470
QY 495 ELGYLVERDRTGDTFRWKGENVSTTEVGTLSRLLDMDADVGVYGVPEVPGTEGRAGMAA 554
D 471 HENFIYFHDVRGDTFRWKGENVATTEVADTVGLVDFOEVNIVGVYVHPDHEGRIGMASIK 530
QY 555 SPTGN--CDLERFAQVLEKELPLYPARPIFLRLPELHKTCTYKFKTELKRGDFPAIVKD 613
D 531 MKENHEFDGKLFQHIADYLPYARPRFLRQDTIETITGTFKHKMTLVEEGFNPVAVIK 590
QY 614 PLFYLDQAKGRVPLDQEAISRIQAGEEKL 643
D 591 ALYFLDDTAKMYVPMTEDIYNAISAKTLKL 620

```

# RESULT 4

```

VLCS_RAT
ID VLCS_RAT STANDARD; PRT; 620 AA.
AC P97524;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE VERY-LONG-CHAIN ACYL-COA SYNTHETASE (EC 6.2.1.-) (VERY-LONG-CHAIN-
DE FATTY-ACID-COA LIGASE).
GN FACVLI OR VLACS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-WISTAR; TISSUE=Liver;
RX MEDLINE=97094763; PubMed=8939997;
RA Uchiyama A., Aoyama T., Kamijo K., Uchida Y., Kondo N., Orii T.,
RA Hashimoto T.;
RT "Molecular cloning of cDNA encoding rat very long-chain acyl-CoA
RT synthetase.";
RL J. Biol. Chem. 271:30360-30365(1996).
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
CC -|- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT

```

## COVALENT BINDING OF AMP TO THEIR SUBSTRATE.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D85100; BAA12722.1; -.
DR InterPro: IPR000873; -.
DR Pfam: PF00501; AMP-binding; 1.
DR PROSITE: PS00455; AMP_BINDING; 1.
KW Ligase; Fatty acid metabolism; Peroxisome; Transmembrane.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 107 127 POTENTIAL.
FT TRANSMEM 262 282 POTENTIAL.
SQ SEQUENCE 620 AA; 70693 MW; 6CF9362DC3805526 CRC64;

```

Query Match 30.9%; Score 1044; DB 1; Length 620;  
Best Local Similarity 38.0%; Pred. No. 4.8e-79;  
Matches 237; Conservative 113; Mismatches 241; Indels 32; Gaps 13;

```

QY 32 LYLGGSGWRIRVFIKT-----IRRDFGLVLLKAKVRQCLQER--RTVPILFASTVRR 85
D 5 LYTGLAGLLPLLLTCCPCPYLLQDVRFVQLANMARQVRSYQRRPVRTILHVFLEQAR 64
QY 86 RHPDKTALIFEGTHTHFERQDEYSSSVANFLOAR--GLASGDVAALFMRNNEFFVGLW 144
D 65 KTHPKPELLFR--DETUTYAQVDRRSNQVARALHDHGLRQGDVALLMGNEPAYVWLW 122
QY 145 GMALGVEAALINTNRDRLHCLTTSRARALVFGESEMAAICEVHASL--DPSLSLFC 202
D 123 GLLKLGCAACLVNIRAKSLLHCFCCGAKVLLVSPQLQAQVEELPSLKDDVYIYVS 182
QY 203 SGSEWEGAVPSTHELDPLKLD--APKHPSCPDGKTDKLFYITSGTGLPKAAIYVHS 261
D 183 SRTSNTNGVDTVLKDVGSADPIPESWRS--EVTFTTTPAVIYITSGTGLPKAAIYVHS 240
QY 262 RYRMAALVYGFMRPNIDIVDCLPLYHSAGNIVGIGOCLLHGMTVIRKKFS 315
D 241 R-----LWGTSLALRSIGIKAHDIYITMPLYHSAALMIGLHGCIVVAGATLARTKFS 293
QY 316 ASRFWDCCIKNCTIVQYIGELCRYLLNQPPREAEHQVRLMALGNLGRSQSTWTFSSRF 375
D 294 ASQFWDCCRKNYNTVQYIGELRLYLCSNPKPDRDHKVRALGNLGRDWRQFVKRF 353
QY 376 HIPOVAEFYATGECNCSLGNEDSOVACGFSNRLSFIYVPIRLVRYNEDTMELIRGPDV 435
D 354 GDHIYEFYATGEGNIFGNFYARKYAGVRVNYLQKKIITLYLDIYDEKDEPVRDENG 413
QY 436 CIPQCPGEGVLGRRIQKDLRRFDGYL--NOGANNKIAKDVKFKGQDQAVLTGDLVMD 494
D 414 CIKVPKGEVGLLVCKITELTP---FFGYAGGKTQTEKKLRDVKFGDLYNSGDLWMD 470
QY 495 ELGYLVERDRTGDTFRWKGENVSTTEVGTLSRLLDMDADVGVYGVPEVPGTEGRAGMAA 554
D 471 RENFIYFHDVRGDTFRWKGENVATTEVADTVGLVDFOEVNIVGVYVHPDHEGRIGMASIK 530
QY 555 SPTG--NCDLERFAQVLEKELPLYPARPIFLRLPELHKTCTYKFKTELKRGDFPAIVKD 613
D 531 MKENYEFNGKLFQHIADYLPYARPRFLRQDTIETITGTFKHKMTLVEEGFNPVAVIK 590
QY 614 PLFYLDQAKGRVPLDQEAISRI 636
D 591 TLYFMDDTKTYVPMTEDIYNAI 613

```

## RESULT 5

VLCS\_MOUSE STANDARD; PRT; 620 AA.

Q35488;  
 15-JUL-1998 (Rel. 36, Created)  
 15-JUL-1998 (Rel. 36, Last sequence update)  
 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE VERY-LONG-CHAIN ACYL-COA SYNTHETASE (EC 6.2.1.-) (VERY-LONG-CHAIN-FATTY-ACID-COA LIGASE).  
 GN FACVL OR VLACS OR VLCS.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Liver;  
 RA Kemp S., Lu J.-F., Smith K.D.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.  
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT COVALENT BINDING OF AMP TO THEIR SUBSTRATE.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF033031; AAB87982.1; -  
 DR InterPro: IPR000873; -  
 DR Pfam: PF00501; AMP-binding; 1.  
 DR PROSITE; PS00455; AMP\_BINDING; 1.  
 KW Ligase; Fatty acid metabolism; Peroxisome; Transmembrane.  
 FT TRANSMEM 1 21 POTENTIAL.  
 FT TRANSMEM 107 127 POTENTIAL.  
 FT TRANSMEM 262 282 POTENTIAL.  
 SQ SEQUENCE 620 AA; 70366 MW; 77C98BD0DE3B9FFB CRC64;

Query Match 30.2%; Score 1022; DB 1; Length 620;  
 Best Local Similarity 37.7%; Pred. No. 3.2e-77;  
 Matches 235; Conservative 108; Mismatches 248; Indels 32; Gaps 12;

QY 32 LYLGGGWRIRVFIKT----IRRDIFGGLVLLKVKAKVQCLOER--RTVPILFASTVR 85  
 Db 5 LYTGLGLLLPLLLTCCCYLLQDVRYFLQLANMARRVRSYQRPRVTLIRAFLEQAR 64  
 QY 86 RHPDKTALIFEQTDTHWTFRLQDEYSSSVANFLQAR--GLASGDVAIAIFMNRNEFYGLWL 144  
 Db 65 KTHPKPELLFR--DETLYAQVDRRSNOVARALHDQLGLRQGDVAFMGNEPAYVWIWL 122  
 QY 145 GMKILGVEAALINLNRDALLCLTTSRARALVFGSEMASACEVHASL--DPSLSLFC 202  
 Db 123 GLLKLGCPMACLNYNIRAKSLHCFOCCGAKVLLASPDLOEAEEALPTLKDAVSFFVY 182  
 QY 203 SGSEPCAVPSPVEHEDPL-LKDAKHLPCSPDKGTFKLFYIYTSGTGLPKAAALVWHS 261  
 Db 183 SRTSNTNGVDTLKVDGVAETPESWRS--EVTETTPAVIYTSGTGLPKAAALINH 240  
 QY 262 RYRMAALVYIGFRM-----RPNDIVDCLPLYSAGNTVIGTQCQLLHGTMVIRKFS 315  
 Db 241 R-----LWYGTGLAMSSGITAQDVITYTTPMLYHSAALMIGLHGICVVGATLARSFS 293  
 QY 316 ASRFWDICIKYNTVIOYIGELCYLLNNOPPREAHNOHVRMALGNLRSQISWTNFSRF 375  
 Db 294 ASQFWDCKRYNTVIOYIGELLYLNCNTPQKPNDRDHVKKALGNLGRDGMVREIKFR 353  
 QY 376 HIPOVAEFGATECNCSLGNFDSQVACGNSRILSFVYPIRLVRYNEDTMELIRPGDV 435  
 Db 354 GDHIVFEYASTEIGNIGFVNYPRKIGAVGRANVYLRQKVARLYELIKYDVKEDEPVDANGY 413  
 QY 436 CIPQCPPEPQLVGRITIQKPLRFDFGYL--NQANKNKTKAKDVFYKGGDQAYLTGDLVMD 494

Db 414 CIKVPKGEVLLVCKITQLTP---FIGVAGKGTOTEKKKLRDVEFKKGDIIYFNSGDLMLID 470  
 QY 495 ELGILYFRDRTGDFRWNKGENVSTTEVEGTLRSRLDMADVAVYGEVPGTEGRAGMAAV- 553  
 Db 471 RENFVYFHDVRDGTFRWNKGENVATTEVADIVGLVDVEEVNNGVYGVPGHEGRIGMASLK 530  
 QY 554 ASPGNCOLERFAQVLEKEPLYPARPIFLRLPLDELHKTGYTKFQKTELRKEGPDPAIVKD 613  
 Db 531 IKENYEFNGKLFQHIAEYLPSPRFLRIQDTIETITGTFKKRKVTLMBEGFNPTVIKD 590  
 QY 614 PLFYLDACKGRYVPLDQEAYSRI 636  
 Db 591 TLYFMDDAEKTVPMTENIYNAI 613  
 RESULT 6  
 ID FAT1\_YEAST STANDARD; PRT; 623 AA.  
 AC P38225;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE PROBABLE LONG-CHAIN FATTY ACID TRANSPORT PROTEIN.  
 GN FAT1 OR YBR041W OR YBR0411.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycetia; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RA Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A.,  
 RA Vissers S.;  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP CHARACTERIZATION.  
 RC STRAIN=W303A;  
 RC MEDLINE-97236810; PubMed-9079682;  
 RA Faergeman N.J., Dirusso C.C., Eiberger A., Knudsen J., Black P.N.;  
 RT "Disruption of the Saccharomyces cerevisiae homologue to the murine fatty acid transport protein impairs uptake and growth on long-chain fatty acids.";  
 RL J. Biol. Chem. 272:8531-8538(1997).  
 CC -1- FUNCTION: MAY BE INVOLVED IN LONG-CHAIN FATTY ACIDS UPTAKE. AND THIS MAY PLAY A PIVOTAL ROLE IN REGULATING THEIR ACCESSIBILITY PRIOR TO METABOLIC UTILIZATION. MAY PLAY AN IMPORTANT ROLE IN UPTAKE OF THESE HYDROPHOBIC COMPOUNDS UNDER CONDITIONS WHERE FATTY ACID SYNTHESIS IS COMPROMISED.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -1- SIMILARITY: TO ENZYMES WHICH ACT VIA AN ATP-DEPENDENT COVALENT BINDING OF AMP TO THEIR SUBSTRATE.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z35910; CA84983.1; -  
 DR PIR; S45899; S45899.  
 DR SGD; S0000245; FAT1.  
 DR InterPro: IPR000873; -  
 DR Pfam: PF00501; AMP-binding; 1.  
 DR PROSITE; PS00455; AMP\_BINDING; 1.  
 KW Lipid transport; Transmembrane.  
 FT TRANSMEM 6 26 POTENTIAL.  
 FT TRANSMEM 54 71 POTENTIAL.  
 FT TRANSMEM 149 169 POTENTIAL.  
 FT TRANSMEM 293 313 POTENTIAL.  
 FT CARBOHYD 184 184 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 289 289 N-LINKED (GLCNAC... ) (POTENTIAL).



FT	CARBOHYD	534	534	N-LINKED (GLCNAC...)	(POTENTIAL).
FT	CARBOHYD	591	591	N-LINKED (GLCNAC...)	(POTENTIAL).
SQ	SEQUENCE	623 AA;	71697 MW;	OAE02700B60C8CFE	CRC64;

  

Query Match	21.1%;	Score 714;	DB 1;	Length 623;
Best Local Similarity	31.8%;	Pred. No. 1.4e-51;		
Matches 205;	Conservative 92;	Mismatches 260;	Indels 88;	Gaps 19;

  

QY	1	MLLGASLVGVLLFSKVLKLPFWQTQVGSLLFL-----	YLGGGWRFRVFRIKT-	48
Db	:	: : : : :	: : : : :	:
Db	7	WFAISRIFFLLPR--LIKLIITPIQKSGLYGNYFDELDRKYRVEDHWIIPYFLKSV	64	
QY	49	-----IRRDI FGLVLVKVAKVRQCQLQRRTVPILFASTVRRHPDKTALIF-----	E	96
Db	:	: : : : :	: : : : :	:
Db	65	FCYIIDVRRHRFNW-----	YLFIKVOQGDLAISYTRPMAEK	104
QY	97	GTDTHTWFRLQDEYS-----SSVANFLQARGLASGDVAATFMENRNFEVGLGMCAKLGV	151	
Db	:	: : : : :	: : : : :	:
Db	105	GEQLETFITYEYNIVLRLSHLHF--DYNVQAGDVADCTNKPFLVFLWSLWNIGA	162	
QY	152	EAMLINTNLRRDALLHCLTTSRARALVFGSEMASACEVHASLDPSLSLCSGSWEPAV	211	
Db	:	: : : : :	: : : : :	:
Db	163	IPAFLANTKGTPLVSLKISNTIQVIDPDASNIRESEEEKNAL-----PDVK	213	
QY	212	PPSTEHLND---PLLKADAPKHLPCDPK---GFTD--KLFYIYTSGTGLPKAAIVVHSR	262	
Db	:	: : : : :	: : : : :	:
Db	214	LNVLEQDLMHLLNSQSPEFLQODNNVTPLGITDFKPSMLIYTSGTGLPKSAIMSWRK	273	
QY	263	YYMAALVYVGFMRPNDIVDCLPLYHSAGNIYGCQCLLHGNTVVIRKKFSASRFWD	322	
Db	:	: : : : :	: : : : :	:
Db	274	SSVGCVQFHVLTWMTNESTVFTAMPLFHSYAALLGACAILLSHGCSLASHKFSASTFWKQ	333	
QY	323	CIRYNTCIYQIGELCYELLNQPPRAENOHOYRMALGNCLRISIWTFNSSRHFIQVAE	382	
Db	:	: : : : :	: : : : :	:
Db	334	VILTGAHQYGVGEVCRIYLHTFISKYEKMHKVAYNGRLRPDIWQDFKRKNIEVIGE	393	
QY	383	FYGATCNCGLNF---DSQVGACGFNSRILSFYPYIR--LVRV--NEDTMELIRPGDV	435	
Db	:	: : : : :	: : : : :	:
Db	394	FYAATEAFATTTFKRGDFGIGACRYNGTIIQWFLSQFTLVRMDPNDSDV-IYRNSKGF	452	
QY	436	CIPCQEPGECOLVGR-IQKDPLRRDGYLNQCANNK-KIAKDVFPKGOAYITGDVLVM	493	
Db	:	: : : : :	: : : : :	:
Db	453	CEVAPVGPGEMLMRLFFPKPTSPQGYLGNNAKETSKVYRDVFRFGDAWYRCGDLKA	512	
QY	494	DELGYLYFRDRTGRFWKGENYSTTEVEGS--TL SRLDDMADVAVYGVPEVPGTEGRAGMA	551	
Db	:	: : : : :	: : : : :	:
Db	513	DEYGLWYFLDRMGDTFRWKSENVSTTEVEDOLTASNKEQYAOVLVIGIKVPKVEGRAGFA	572	
QY	552	AVASPNGCDLERFAOVLEK-----ELPIYARPIFLRLPELHKT	591	
Db	:	: : : : :	: : : : :	:
Db	573	VIKLTNSLDITAKTKLLNDSLRLNPYAMPFLFYKFVDIKMT	617	

  

RESULT	7
CAIC.ECOLI	
ID	CAIC.ECOLI STANDARD; PRT; 522 AA.
AD	P31552;
DT	01-JUL-1993 (Rel. 26, Created)
DT	01-JUL-1993 (Rel. 26, Last sequence update)
DT	15-DEC-1998 (Rel. 37, Last annotation update)
DE	PROBABLE CROTONOBETAINE/CARNITINE-COA LIGASE (EC 6.3.2.-).
DE	CAIC.
OS	Escherichia coli.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC	Escherichia.
OX	NCBI_TaxID=562;
RN	[1]
RRP	SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC	STRAIN=O44 K74;
RX	MEDLINE=95115548; PubMed=7815937;
RA	Eichler K., Bourgis F., Buchet A., Kleber H.-P., Mandrand-Berthelot M.-A.;

"Molecular characterization of the *cat* operon necessary for carnitine metabolism in *Escherichia coli*."; Mol. Microbiol. 13:775-786(1994).  
[2]  
SEQUENCE FROM N.A.  
RN STRAIN-K12.  
RC MEDLINE=92334977; PubMed=1630901;  
RX MEDLINE=97426637; PubMed=9278503;  
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,  
RA Isono K., Mizobuchi K., Nakata A.;  
RA "Systematic sequencing of the *Escherichia coli* genome: analysis of  
RN the 0-2.4 min region."; Nucleic Acids Res. 20:3305-3308(1992).  
RN [3]  
SEQUENCE FROM N.A.  
RN STRAIN-K12 / MG1655;  
RC MEDLINE=97426637; PubMed=9278503;  
RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RN "The complete genome sequence of *Escherichia coli* K-12."; Science 277:1453-1474(1997).  
RL  
CC -1- FUNCTION: COULD CATALYZES THE TRANSFER OF COA TO CROTONOBETAINE  
CC OR CARNITINE.  
CC  
CC -1- PATHWAY: CARNITINE METABOLISM.  
CC  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
CC (POTENTIAL).  
CC  
CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT  
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

Query Match      10.4%; Score 350.5; DB 1; Length 522;
Best Local Similarity 24.1%; Pred. No. 2e-21;
Matches 137; Conservative 86; Mismatches 275; Indels 71; Gaps 18;

OY 52 DIFGLVLKLVKAVRQCLQERTVPILEASTVRRHPDKTALIFE---GTDTHWTFOLD 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7 DIIG-----QHRLQWMDLAD-----YGHKTALICSSGGVWNRYSYELN 49

```

QY 109 EYSSVANFLOARGLASGDVAIAFEMENREFFVGLWMAKLGVEAALINTNLRDALLHC 168  
Db 50 QEINRTANFYTLGIRKGDVALHLDNCPFIICFPGAKIGAIMVPIINARLCEESAWI 109  
QY 169 LTTSRARALVFGSEMASACEVHASLDPSSLFCSGSWEPGAVP--PSTEHLDPLLKQAP 226  
Db 110 LQNSQACLLVTSQAQFPMYQQIQEEDATQRLHIC---LTDVALPADDGVSSFTQLKNQOP 166  
QY 227 KHLPCSPDKGFTDKLFYIYTSQGLPKAAIVVHSRYRYMAALVYGG---FMRPNDIVY 283  
Db 167 ATLCYAPPLSTDDTAEILFTSGTTSRPGKWITH---YNLRFAGYYSAWOQALRDDVYL 223  
QY 284 DCLPLVHSAGNIVGICQILLHGMVTVIRKFSASRFWDCCIYKYNCTIVQYIGELCRYLLN 343  
Db 224 TVMPAFHIDCQCTAAMAAFSAGATFVLVEKYSARAFWQGVQYRATVTECIPMTIRLMV 283  
QY 344 OPPEAENHOVRMALGN--GLRQSIWTFNFSRFHPOVAEFGYATGECNCS--LGNEFDS--- 398  
Db 284 QPSANDQOHLREVMFYLNLSQEKDAFCERFV--RLTSGYMTETIVGIIIGDRPGDKR 342  
QY 399 ---OVGACGFSRLSFVYPIRLVRVNDTLMELIRPGD--VCIPQCPGPGOLVRIQK 454  
Db 343 RWPISGRVGF-----CYEAER---DDHNRPLPAGEIGEICI---KGIPGKTIK--- 386  
QY 455 DPLRRFDGVLNQGANNKKTAKDKVFKGQAYL--TGDVLYMDELGYLYFDRDTGDTFRWKG 513  
Db 387 -----EYLNPOATAKVLEAD-----GWLHTGDTGTRDEEDFFYVDRRCNMIRKGG 433  
QY 514 ENVSTTEVEGTLRLDMDADVAVYGVVEPTEGRAGMAAASVPTGNCDLREFAQVLEKEL 573  
Db 434 ENVSCVELENIAAHPKIQDVIIVVGKIDSRDEAIKAFVVLNEGETLSSEEFRCQNM 493  
QY 574 PLYARPIFLRLPELHKHTGTGYFQKTEL 602  
Db 494 AKFKVPSYLEIRKDLPRNGSGKIIRKNLK 522

RESULT 8  
ACSA\_COPCI STANDARD; PRT; 661 AA.  
AC 013440;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE ACETYL-COENZYME A SYNTHETASE (EC 6.2.1.1) (ACETATE--COA LIGASE) (ACYL-  
DE ACTIVATING ENZYME).  
GN ACS-1.  
OS Coprinus cinereus (Inky cap fungus).  
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Agaricales;  
OC Coprinaceae; Coprinus.  
OX NCBI\_TaxID=5346;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-JV6;  
RA Chauré P.T., Casselton L.A., Connerton I.F.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: ATP + ACETATE + COA -> AMP + DIPHOSPHATE +  
CC ACETYL-COA.  
CC -!- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT  
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL; Y15417; CAA75612.1; -.  
DR InterPro; IPR000873; -.  
DR Pfam; PF00501; AMP-binding; 1.

DR PRINTS: PR00154; AMP BINDING.  
DR PROSITE; PS00455; AMP\_BINDING; 1.  
KW Ligase.  
SQ SEQUENCE 661 AA; 73094 MW; D647C974095795D9 CRC64;

Query Match 9.4%; Score 318; DB 1; Length 661;  
Best Local Similarity 25.3%; Pred. No. 1.5e-18;  
Matches 146; Conservative 90; Mismatches 240; Indels 100; Gaps 25;  
QY 86 RHPDKALIFE---GTDTHWTFQLDDEYSSVANFLOARGLASGDVAIAFEMENREFFV 141  
Db 96 KHPEKTAIIYEADPEGGEISYAEALLREVCSIANVLKSGVKKGDVSVYLPMTQAVA 155  
QY 142 LWLGMALGVEAALINTNLRDALLHCLTTSRARALVFGSE-----MASAICE 189  
Db 156 AFLACARIGAIHVSVFAGFAEALRDMQCKSRVLTSDEGRGKAIATKAIDAALK 215  
QY 190 VHASLDPSSLFCSGSWEPGAVPSTEHLDPLLDKAPHLPS--CPDK--GFTDKLFVIY 246  
Db 216 ECPAVEKVLVLRKTN-----PVPWTEGRDWWHEAVRVPYCPPEVMASEDPLFIY 270  
QY 247 SGTGLPKAAIVVHSRYRYMAAL--VYGFMRPNDIVDCLPLYHSAGNIVG-----IGQ 300  
Db 271 SGSTGKPGVVHTTGGYLLCAALTVKYVDFVHPDD--RFACMA---DVGWITGHTYIVGP 326  
QY 301 CLLHGMVWIRKK---FSASRFWDCCIYKYNCTIVQYIGELCRYLLNOPPREAENQH--- 353  
Db 327 LAIGATTVFESTPVYPTSPRYMETVEKYLTQF--YSAPTALRLRLRGHEHVNKHDLSS 385  
QY 354 -QVRMALGNLGRSIWTFNFSRFHPOVAEFGYATGECNCSLGNFDSQVAGCGFNSRLSF 412  
Db 386 LRVLSGVGEPIINPAWHYN-----EHVKTECAIVDTFWQETGS-----IVVT 430  
QY 413 VYPIRLVRVNDTLMELIRPGDVCIPQCPGPGOL-----VGRIOKDLRPRDGYLNOGAN 469  
Db 431 PFGAI-----ETK-----PGAATVPFEGIEPAILETTGKVLGENDV---EGVLTIAHP 477  
QY 470 NKKTAKDVFKGD--QAYL-----TGDLVLYMDELGYLYFDRDTGDTFRWKGENV 516  
Db 478 WPSARTIY---GDHQRYLEYMKPYGYFTTGGGAARDEGDIWIKGRVDDVINVSGHRL 535  
QY 517 STTEVEGTLRLDMDADVAVYGVVEPTEGRAGMAAAYA--SPTGNCDLREFAQVLEKEL 575  
Db 536 STAEISALITHTGVAETAVIGTADLT--GQAVYAVFTLKPPEKFEADENAG--LSKELIL 593  
QY 576 -----YARPIFLRLPELHKHTGTGYFQKTELK 603  
Db 594 QVRKIIGPFAAPKRIYIVSDLPKTRSGKIMRILRK 629

RESULT 9  
LCFA\_ECOLI STANDARD; PRT; 561 AA.  
ID LCFA\_ECOLI  
AC P29212;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE LONG-CHAIN-FATTY-ACID--COA LIGASE (EC 6.2.1.3) (LONG-CHAIN ACYL-COA  
DE SYNTHETASE).  
DE FADD OR OLD.  
GN Escherichia coli.  
OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia coli.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE=941150456; PubMed=8107670;  
RA Fulda M., Heinz E., Wolter F.P.;  
RT "The fadD gene of Escherichia coli K12 is located close to rnd at  
RT 39.6 min of the chromosomal map and is a new member of the  
RT AMP-binding protein family.";



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC The European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M62755; AAA33842.1; -  
 DR PIR: A39827; A39827.  
 DR HSSP: P08659; ILIC1.  
 DR InterPro: IPR000873; -  
 DR Pfam: PF00501; AMP-binding; 1.  
 DR PROSITE: PS00455; AMP-BINDING; 1.  
 KW Ligase; phenylpropanoid metabolism; Multigene family.  
 SQ SEQUENCE 545 AA; 59619 MW; DE183683B774BA71 CRC64;

Query Match 9.0%; Score 303; DB 1; Length 545;  
 Best Local Similarity 24.0%; Pred. No. 2e-17;  
 Matches 131; Conservative 98; Mismatches 231; Indels 86; Gaps 22;

QY 90 KTALEFGTDTHTFQRLDEYSSVANFLQARGLASGDVAALFMENRNEFVGLWGLMAKL 149  
 DB 44 RPLCLDGANDRIYTYAEVELTSRKVAVGLNKLGIQKQDTIMILLPNCPEFVFAFTGASYL 103  
 QY 150 GVEAALINTNLRDALLHCLTTSRRAALVFGSEMASAICEVHASLDPSLSLFCSGSWEPG 209  
 DB 104 GAISTMANPLFTPAEVVQAKASSAKIVITQACFAGKVQDY--AIENDLKVICVDS---- 157  
 QY 210 AVPPSTHELDPLKDAKHLPSG---PDGFTDKLFYIYTSGTGLPKAAIVVHRSRYRM 266  
 DB 158 -VPEGCVHSELQSDHEIHPVKIQPD---DVVALPYSSGTTGLPKGVMTHKGLVTS 212  
 QY 267 AALVYVG---FMRPNDIVDCLPLYHSAGNIVGIGQCLL---HGMTVVRKKFSASR 318  
 DB 213 VAQOVGDNANLYMHSDVLMCVLPFH---IYLSNVLLCALRVGAAILIMQKFDIAQ 268  
 QY 319 FWDDCIKYNTVOYIGELCRYLLNQPPREAENQHVRMA-----LGNGLRQSIWTFNS 372  
 DB 269 FLELIPKHVKTIGFPVPIVLAIAKSPVDVNDLSSVTVMSGAAPLKGLEDAV---R 324  
 QY 373 SRFHIPQVAEYFATGECNCSLG-----NFDQVAGCGFNSRILSFYPIRLVRVNE 424  
 DB 325 AKFPNAKLGQGYGTGAGPVLMCLAFAPKPEPDIKSGAGC-----TVVRNAB- 371  
 QY 425 TMELIRPGDVCIPGCPGPGQVGVRIQKDLPLRRFDGVLNCGANNKKIAKDVFRKQDOA 484  
 DB 372 -MKIVDPDTGCSLP--RNQPGEI--CIRGDIQIMK--GYLN--DPEATARTIEKEG--- 417  
 QY 485 YL-TGDLVLMDELGYLYFRDRTGDTFRWKGENVSTTEVEGTLRSLDMDADVAVGYVEVP 543  
 DB 418 WLHTGIDGIFDIDDDDELFIVDRLKELIKYKGFQVAPAELEALLINHPDISDAV---VPM 473  
 QY 544 TEGRAGMAVA---SPTGNCDLERFAQVLEKELPLY---ARPIFLRLLPELHKTKTYKF 596  
 DB 474 IDEQAGEVPVAVVRNSNGSTITEDEVKDFISKQVIFYKRIKRVFFVETVP---KSPSGKI 530  
 QY 597 QKTEL 602  
 DB 531 LRKDLR 536

RESULT 11  
 ID 4CL2\_SOLTU STANDARD; PRT; 545 AA.  
 AC P31685;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE 4-COUMARATE--COA LIGASE 2 (EC 6.2.1.12) (4CL 2) (4-COUMAROYL-COA  
 DE SYNTHASE 2).

GN 4CL2 OR 4CL-2.  
 OS Solanum tuberosum (Potato).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
 OC Solanales; Solanaceae; Solanum.  
 OX NCBI\_TaxID=4113;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91217100; PubMed=2022667;  
 RA Becker-Andre M., Schulze-Lefert P., Hahlbrock K.;  
 RT "Structural comparison, modes of expression, and putative cis-acting  
 RT elements of the two 4-coumarate: CoA ligase genes in potato.";  
 RL J. Biol. Chem. 266:8551-8559(1991).  
 CC -1- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE +  
 CC 4-COUMAROYL-COA.  
 CC -1- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID  
 CC METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.  
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT  
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.  
 DR PIR: B39827; B39827.  
 DR HSSP: P08659; ILIC1.  
 DR InterPro: IPR000873; -  
 DR Pfam: PF00501; AMP-binding; 1.  
 DR PROSITE: PS00455; AMP-BINDING; 1.  
 KW Ligase; Phenylpropanoid metabolism; Multigene family.  
 SQ SEQUENCE 545 AA; 59625 MW; 5481F0B0AFA39E0 CRC64;

Query Match 8.9%; Score 300; DB 1; Length 545;  
 Best Local Similarity 24.1%; Pred. No. 3.5e-17;  
 Matches 132; Conservative 96; Mismatches 230; Indels 90; Gaps 21;

QY 90 KTALEFGTDTHTFQRLDEYSSVANFLQARGLASGDVAALFMENRNEFVGLWGLMAKL 149  
 DB 44 RPLCLDGANDRIYTYAEVELTSRKVAVGLNKLGIQKQDTIMILLPNCPEFVFAFTGASYL 103  
 QY 150 GVEAALINTNLRDALLHCLTTSRRAALVFGSEMASAICEVHASLDPSLSLFCSGSWEPG 209  
 DB 104 GAISTMANPLFTPAEVVQAKASSAKIVITQACFAGKVQDY--AIENDLKVICVDSAPG 161  
 QY 210 AVP-----PSTHELDPLKDAKHLPSGPKGFTDKLFYIYTSGTGLPKAAIVVHRSRY 264  
 DB 162 CVHSELQSDHEIHPVKIQPDVWALP-----YSSGTTGLPKGVMTHKGLV 210  
 QY 265 RMAALVYVG---FMRPNDIVDCLPLYHSAGNIVGIGQCLL---HGMTVVRKKFSA 316  
 DB 211 TSAQOVGDNANLYMHSDVLMCVLPFH---IYLSNVLLCALRVGAAILIMQKFDI 266  
 QY 317 SRFWDDCIKYNCTVOYIGELCRYLLNQPPREAENQHVRMA-----LGNGLRQSIWTFN 370  
 DB 267 AQFLELIPKHVKTIGFPVPIVLAIAKSPVDVNDLSSVTVMSGAAPLKGLEDAV--- 323  
 QY 371 FSSRHIPQVAEYFATGECNCSLG-----NFDQVAGCGFNSRILSFYPIRLVRVN 422  
 DB 324 -RAKFPNAKLGQGYGTGAGPVLMCLAFAPKPEPDIKSGAGC-----TVVRNA 370  
 QY 423 EDTMELIRPGDVCIPGCPGPGQVGVRIQKDLPLRRFDGVLNCGANNKKIAKDVKKGD 482  
 DB 371 E--MKIVDPDTGCSLP--RNQPGEI--CIRGDIQIMK--GYLN--DPEATARTIEKEG- 417  
 QY 483 QAYL-TGDLVLMDELGYLYFRDRTGDTFRWKGENVSTTEVEGTLRSLDMDADVAVGYVE 541  
 DB 418 --WLHTGIDGIFDIDDDDELFIVDRLKELIKYKGFQVAPAELEALLINHPDISDAV---V 471  
 QY 542 PGTTEGRAGMAVA---SPTGNCDLERFAQVLEKELPLY---ARPIFLRLLPELHKTKTY 594  
 DB 472 PMIDEQAGEVPVAVVRNSNGSTITEDEVKDFISKQVIFYKRIKRVFFVETVP---KSPSG 528  
 QY 595 KFKTEL 602  
 DB 529 KILKDLR 536

RESULT 12  
4CL1\_TOBAC

ID 4CL1\_TOBAC STANDARD; PRT; 547 AA.

AC 024145; PRT; 547 AA.

DT 01-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE 4-COUMARATE--COA LIGASE 1 (EC 6.2.1.12) (4CL 1) (4-COUMAROYL-COA

DE SYNTHASE 1).

GN 4CL1

OS Nicotiana tabacum (Common tobacco).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;

OC Solanales; Solanaceae; Nicotiana.

OX NCBI\_TaxID=4097;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96416441; PubMed=8819324;

RA Lee D., Douglas C.J.;

RT "Two divergent members of a tobacco 4-coumarate:coenzyme A ligase

RT (4CL) gene family. cDNA structure, gene inheritance and expression,

RT and properties of recombinant proteins.";

RL Plant Physiol. 112:193-205(1996).

CC -|- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE +

CC 4-COUMAROYL-COA.

CC -|- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID

CC METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.

CC -|- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT

CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC EMBL; U50845; AAB18637.1; .

CC DR HSP; P08659; 1LCI.

DR Mendel; 24477; Nicta; 1179; 24477.

DR InterPro; IPR000873; .

DR Pfam; PF00501; AMP-binding; 1.

DR PROSITE; PS00455; AMP\_BINDING; 1.

DR Ligase; Phenylpropanoid metabolism; Multigene family.

SK SEQUENCE 547 AA; 59842 MW; 154DF6D684E3F51B CRC64;

SEQUENCE

Query Match 8.8%; Score 299; DB 1; Length 547;

Best Local Similarity 24.0%; Pred. No. 4.2e-17;

Matches 132; Conservative 93; Mismatches 232; Indels 92; Gaps 22;

QY 90 KTALEFGDTHTFRLQDEYSSSVANFLQARGLASGDVAIFENRNEFFVGLWGLMAKL 149

Db 46 RCLINGANDQIYTAELVETCRKAVGLNKLIGIOQKDTIMLLPNSEFFVFAFMGASYL 105

QY 150 GVEAALINTLRDALLHCLTTSRARALVFGSEMAISACEVHASLDPSLSLFCSGSEWPG 209

Db 106 GAISTMANPLETPAEVWVKQAKASSAKIITQSCFVGKVDYASEND--VKVICIDSAPEG 163

QY 210 A-----VPPSTHLDPLKADKHPKLPSCDPKGFDTKLYIYTSGTGLPKAAIVVHSRY 264

Db 164 CLHFSELQSDSEIPEVKIQDDVVALLP-----YSSGTGLPKGVMLTHGLV 212

QY 265 RMAALVYTG-----FRMRPNIDVDCPLPLHSAGNIVGICOLLHGM-----TWVIRKFFSA 316

Db 213 TSVAQQVDGENANLYMHSEDLVLCVLPFLH---IYLSNILLGLRVAAILIMQKFDI 268

QY 317 SRFWDCCIYKNTVOYIGELCRYLLNQPPEAENQHVRMA-----LGNLRQSIWTN 370

Db 269 APFLEIQKYKVSFPFPVPLAIAKSPIDVDSVTVMSGAAPLGKELEDAYRTK 328

QY 371 FSSRHIFQVAFFYGATCNCSLG-----NFDQVGACGCFNSRILSFVPIRLVRVN 422

Db 329 FPN-----AKLGGYGMTEAGPVLAMCLAFAPKPEFDIKSGACG-----TVVRNA 372

QY 423 EDTMELIRGPGVCIPCPQPGPGLVGRILQKDLRRFDGYLNQ--GANNKKIADKPFKK 481

Db 373 E--MKIVDPDTGCSLP--RNQFGEI---CIRGDQIMK--GYLNDPEATRTFRIDK----- 417

QY 482 DOAYL-TGDVLVMDLGYLYFRDRTGDFRWKGENVSTVEGTLSRLLDMAVAVYGV 540

Db 418 -EGWLHTGDIGFIDEDELFTVDRKLKELIKYKGFQVAPAEATEALLNHPNISDAV----- 472

QY 541 VPGTEGRAG---MAAVASPTGNCNLE-RFAQVLEKELPLY---ARPLRLLLPELHGTGT 593

Db 473 VPMKDEQAGEVPFVAFVVRKNSGSAITEDEVKDFISKVIFYKRVRRVFEVETVP---KSPS 529

QY 594 YKFORTELR 602

Db 530 GKILRKDLR 538

RESULT 13

4CL3\_ARATH

ID 4CL3\_ARATH STANDARD; PRT; 561 AA.

AC 095777;

DT 01-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE 4-COUMARATE--COA LIGASE 3 (EC 6.2.1.12) (4CL 3) (4-COUMAROYL-COA

DE SYNTHASE 3).

GN 4CL3 OR AT1G65060 OR F16G16.6.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

OC Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=CV. COLUMBIA;

RX MEDLINE=99348176; PubMed=10417722;

RA Ehling J., Buettner D., Wang Q., Douglas C.J., Somssich I.E.,

RA Kombrink E.;

RT "Three 4-coumarate:coenzyme A ligases in Arabidopsis thaliana

RT represent two evolutionarily divergent classes in angiosperms.";

RL Plant J. 19:9-20(1999).

RN [2]

RP SEQUENCE FROM N.A.

RX STRAIN=CV. COLUMBIA;

RX MEDLINE=21016719; PubMed=11130712;

RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,

RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,

RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,

RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,

RA Dunn P., Egu P., Feldlyum T.V., Feng J.-D., Fong B., Fujii C.Y.,

RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,

RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,

RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,

RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,

RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A.,

RA Millschler J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,

RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,

RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,

RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,

RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,

RA Wu D., Xu G., Fraser C.M., Venter J.C., Davis R.W.;

RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis

RT thaliana.";

RL Nature 408:816-820(2000).

CC -|- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE +

CC 4-COUMAROYL-COA.

CC -|- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID

CC METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.

CC -|- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT

CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: AF106087; RAD47194.1; -;  
 DR EMBL: AF106088; RAD47195.1; -;  
 DR EMBL: AC009360; AAF06039.1; -;  
 DR HSSP: P08659; ILCI.  
 DR InterPro: IPR000873; -;  
 DR Pfam: PF00501; AMP-binding; 1.  
 DR PROSITE: PS00455; AMP-BINDING; 1.  
 KW Ligase; Phenylpropanoid metabolism; Multigene family.  
 SQ SEQUENCE 561 AA; 61310 MW; C2EFFIC36B33F6DC CRC64;

Query Match 8.7%; Score 293.5; DB 1; Length 561;  
 Best Local Similarity 25.3%; Pred. No. 1.3e-16;  
 Matches 139; Conservative 84; Mismatches 237; Indels 89; Gaps 23;

QY 89 DKTALIFEGTDTHTWTRQLEDEYSSVANFLQARGLASGDVAIAFMENRNEFVGLMLGMAK 148  
 DB 63 DKPCLLIVGSTGKSYTYGETHLICRRVASGLYKLRKGDVIMILLQNSAEFFSEFMGASM 122  
 QY 149 LGVEAALINTNRDALLHCLTTSRRALVFGSEMAAICEVHASLDPSLSFCGSGWEP 208  
 DB 123 IGAVSTTANPFYTSQELYKQLSSGAKLIITHSQYVDKL---KNLGENLTITITDEPTP 178  
 QY 209 -GAVPSTPEHLDPKLPSCDCKGFTDKLYIYTGTTGLPKAAIVVHSRYRMA 267  
 DB 179 ENCLPSTLTID---DETNPFOETVDIGDDAAALPFSSTGTLGPKGVLPKSLITSV 234  
 QY 268 ALVYVG---FRMRPNDIVDCLPLYSAGNIVIGQCILH---GMTVVRKKSASRF 319  
 DB 235 AQVDGDNPNLYLKSNDVILCVLPFH---IYSLNSVLLLSRSGATVLLMKFEIGAL 290  
 QY 320 WDCIKYNTCTIVYIGELCYRLNQPPREANQHQVRMA-----LGNLQRSQIWNFESS 373  
 DB 291 LDLIQRHRTVIAALVPLVIALAKNTVNSYDLSSRVFLSGAALGKELQDSL----- 344  
 QY 374 RFHPO---VAEFTGATEC-----NCSLG---NFDQVAGACGNSRLISVYPIRLVRVNE 423  
 DB 345 RRRLPQAILGGQGMTEAGPVLMSLGRKEPIPTKSGSGC-----TVVRNAELKVVHL 398  
 QY 424 DTWELIRG---PDGVCPICQPGEPGLVGRIOKDLRRFDGLNQ-GANNKKIAKDVFK 479  
 DB 399 ET-RLSLGYNQPGECIR-----GQIMKE-----YLNDEPATSATI----- 434  
 QY 480 KGDQAYL-TGDVLVMDLGLYFRDRTGDTFRWKGENVSTTEVEGTLRLDMDADVYVG 538  
 DB 435 -DEEGWLTGDCYVDEDEIFIVDLKKEVIFKFGQVPPAELESLLINHSDIAA-- 491  
 QY 539 VEPGTEGRAG---MAVASPTGNDL---ERFAQVLEKELPLYPARIFIRLLPELHKGT 593  
 DB 492 --VPQNDVAGEPVFAVVRNSGN-DITEDVKEYVAKQVVFYKRLHKVFFVASIPKSPS 548  
 QY 594 YKFKTELR 602  
 DB 549 GKILRKDLK 557

RESULT 14  
 4CL1\_PETCR  
 ID 4CL1\_PETCR STANDARD; PRT; 544 AA.  
 AC P14912;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE 4-COUMARATE--COA LIGASE 1 (BC 6.2.1.12) (4CL 1) (4-COUMAROYL-COA

DE SYNTHASE 1).  
 GN 4CL1 OR 4CL1-1.  
 OS Petroselinum crispum (Parsley) (Petroselinum hortense).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae;  
 OC euasterids II; Apiales; Apiaceae; Petroselinum.  
 ON NCBI\_TaxID=4043;  
 RX SEQUENCE FROM N.A.  
 RP MEDLINE=89005119; PubMed=3169018;  
 RA Lozoya E., Hoffmann H., Douglas C., Schulz W., Scheel D.,  
 RA Hahlbrock K.;  
 RT "Primary structures and catalytic properties of isoenzymes encoded by  
 RT the two 4-coumarate:CoA ligase genes in parsley.";  
 RL Eur. J. Biochem. 176:661-667(1988).  
 RN [2]  
 RP SEQUENCE OF 1-8 FROM N.A.  
 RA Douglas C., Hoffmann H., Schulz W., Hahlbrock K.;  
 RT "Structure and elicitor or U.V.-light-stimulated expression of two  
 RT 4-coumarate:CoA ligase genes in parsley.";  
 RL EMBO J. 6:1189-1195(1987).  
 CC -1- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE +  
 CC 4-COUMAROYL-COA.  
 CC -1- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID  
 CC METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.  
 CC -1- INDUCTION: TRANSCRIPTIONALLY ACTIVATED BY BOTH FUNGAL ELICITOR AND  
 CC ULTRAVIOLET IRRADIATION.  
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT  
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

EMBL: X13324; CAA31696.1; -;  
 EMBL: X05350; CAA28959.1; -;  
 DR PIR: S01667; S01667.  
 DR HSSP: P08659; ILCI.  
 DR InterPro: IPR000873; -;  
 DR Pfam: PF00501; AMP-binding; 1.  
 DR PROSITE: PS00455; AMP-BINDING; 1.  
 KW Ligase; Phenylpropanoid metabolism; Multigene family.  
 SQ SEQUENCE 544 AA; 59825 MW; 22BBAD78F25D0C8 CRC64;

Query Match 8.5%; Score 286; DB 1; Length 544;  
 Best Local Similarity 22.7%; Pred. No. 5.1e-16;  
 Matches 123; Conservative 99; Mismatches 243; Indels 78; Gaps 20;  
 QY 89 DKTALIFEGTDTHTWTRQLEDEYSSVANFLQARGLASGDVAIAFMENRNEFVGLMLGMAK 148  
 DB 41 DKSLINGATGETTYTSQVELLSRKVASGLNKLGIQGGDTIMLLLPNSPEFFAFLGASY 100  
 QY 149 LGVEAALINTNRDALLHCLTTSRRALVFGSEMAAICEVHA-----SLDPSLSLFCSG 204  
 DB 101 RGAISTMANPFTSAEVIKQLKASQAKLII-----TQACVVDKVKDYAAEKNIQIICID 154  
 QY 205 SWEPCAVPPSTPEHLDPKLPSCDCKGFTDKLYIYTGTTGLPKAAIVVHSRYR 264  
 DB 155 D-----APQDCFLHFSKLEADESEMEPEVINS--DDVVALPYSSGTTGLPKGVMLTHKGLV 208  
 QY 265 RMAALVYVG---FRMRPNDIVDCLPLH---SAGNIVIGQCILHGMTVVRKKSASR 318  
 DB 209 TSAQVQDGDNDPNLMHSEDVWICILPLFIHLSNAVLCCG--LRAGVTILLIMQKFDIVP 266  
 QY 319 FWDCCIKNCTIVYIGELCYRLNQPPREANQHQVRMA-----LGNLQRSQIWNFNS 372  
 DB 267 FLELIQKYKVTIGPVPPIVLAIAKSPVVDKYLSSVRTVMSGAPLGLKLEDAV-----R 322

QY 373 SRPHIQAEEFYGATECNCSLG-----NFSQVQAGCFNSRILSFVYPIRLVRVNE 424  
 Db 323 AKFPNAKLGCGTGTAGVPLAMCIAFAKEPYEIKSGAG-----TIVRNAE- 369  
 QY 425 TMLIRGPDGVCIPQCPGPGGLVGRRIQKDLRRFDGYLNOGANNKKIAKDVFKKQDOA 484  
 Db 370 -MKIVDPETNASLP--RNQGEI--CIRGDIK--GYLNDPESTRTI-----DEG 415  
 QY 485 YL-TGDLVMDDELGYLYFRDRGTDFRKNGENVSTVEGTLSRLDMADVAVYGEVPG 543  
 Db 416 WLHTGIDGIDDDDELFIYDLRLKEIKYKGFQVAPAELEALLTHPTISDAV-----VPM 471  
 QY 544 TEGRAG---MAAVASPTG-NCDLERFAQVLEKELPLYPARLPLPELHKTGTYKFKT 599  
 Db 472 IDEKAGEVPVAVVRNNGFTTTEEEKQFVSKQVYKIRIFRVFFVDAIPKSPSGKILRK 531  
 QY 600 ELR 602  
 Db 532 DLR 534

## RESULT 15

4CL2\_PETCR  
 ID 4CL2\_PETCR STANDARD; PRT; 544 AA.  
 AC P14913;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE 4-COUMARATE-COA LIGASE 1 (EC 6.2.1.12) (4CL 1) (4-COUMAROYL-COA  
 DE SYNTHASE 1).  
 GN 4CL2 OR 4CL2-2.  
 OS Petroselinum crispum (Parsley) (Petroselinum hortense).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Asterales;  
 OC euasterids II; Apiales; Apiaceae; Petroselinum.  
 OX NCBI\_taxid=4043;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89005119; PubMed=3169018;  
 RA Lozoya E., Hoffmann H., Douglas C., Schulz W., Scheel D.,  
 RA Hahlbrock K.;  
 RT "Primary structures and catalytic properties of isoenzymes encoded by  
 RT the two 4-coumarate:COA ligase genes in parsley.";  
 RL Eur. J. Biochem. 176:661-667(1988).  
 RN [2]  
 RP SEQUENCE OF 1-8 FROM N.A.  
 RA Douglas C., Hoffmann H., Schulz W., Hahlbrock K.;  
 RT "Structure and elicitor or U.V.-light-stimulated expression of two  
 RT 4-coumarate:COA ligase genes in parsley.";  
 RL EMBO J. 6:1189-1195(1987).  
 CC -1- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE +  
 CC 4-COUMAROYL-COA.  
 CC -1- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID  
 CC METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.  
 CC -1- INDUCTION: TRANSCRIPTIONALLY ACTIVATED BY BOTH FUNGAL ELICITOR AND  
 CC ULTRAVIOLET IRRADIATION.  
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT  
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X13325; CAA31697.1; -;  
 DR EMBL; X05351; CAA28960.1; -;  
 DR PIR; S01667; S01667.  
 DR PIR; S15695; S15695.  
 DR HSSP; P08659; ILCI.  
 DR InterPro; IPR000873; -;

DR Pfam; PF00501; AMP-binding; 1.  
 DR PROSITE; PS00455; AMP\_BINDING; 1..  
 KW Ligase: Phenylpropanoid metabolism; Multigene family.  
 SQ SEQUENCE 544 AA; 59783 MW; B477965C68F8C534 CRC64;

Query Match 8.3%; Score 282; DB 1; Length 544;  
 Best Local Similarity 22.7%; Pred. No. 1.1e-15;

Matches 123; Conservative 98; Mismatches 244; Indels 78; Gaps 20;

QY 89 DKTALEIEGDTHTTFFQOLDEYSSVANFLOARGLASGDVAAIFMENRNFVGLHGLMAK 148  
 Db 41 DKSLINGATGETFTYSQVELLSKRVASGLKLGIOQGDITMLLLPNSPEYFFAFLGASY 100  
 QY 149 LGVEAALINTLRDALLHCLITTSRRLVFGSEMSAICEVHA-----SLDPSLSLFCSG 204  
 Db 101 RGAISTWANFFTSAEVIKOLKASLAKLII-----TQACYVDKVDYAAEKNIQICID 154  
 QY 205 SWEPGAVPPSTEHLDPKADPKHLPSCDPKGTDKLFYIYTSGTGLPKAAIVVHSRY 264  
 Db 155 D-----APQDCLHFKLMEADESEMPEVVIDS--DDVVALPYSSGTTGLPKGVMLTHKGLV 208  
 QY 265 RMAALVYVG----FRMRPNDIVVDCLPLXH--SAGNIVGIGQCLLHGMTVVIKKFSASR 318  
 Db 209 TSVAQQVDGDNPNLYMHSEDMICILPLFIYSLNAVLCG--LRAGVTILIMQKFDIVP 266  
 QY 319 FWDDCIKYNCTIVQYIGELCRYLLNQPPPPAEANOHOVRMA-----LGNGLRQSIWTFNS 372  
 Db 267 FLELIQYKVTIGPFVFPPIVLAIAKSPVDKYDLSSVRTVMGGAAPLGKLEDAV---R 322  
 QY 373 SRPHIQAEEFYGATECNCSLG-----NFSQVQAGCFNSRILSFVYPIRLVRVNE 424  
 Db 323 AKFPNAKLGCGTGTAGVPLAMCIAFAKEPYEIKSGAG-----TIVRNAE- 369  
 QY 425 TMLIRGPDGVCIPQCPGPGGLVGRRIQKDLRRFDGYLNOGANNKKIAKDVFKKQDOA 484  
 Db 370 -MKIVDPETNASLP--RNQGEI--CIRGDIK--GYLNDPESTRTI-----DEG 415  
 QY 485 YL-TGDLVMDDELGYLYFRDRGTDFRKNGENVSTVEGTLSRLDMADVAVYGEVPG 543  
 Db 416 WLHTGIDGIDDDDELFIYDLRLKEIKYKGFQVAPAELEALLTHPTISDAV-----VPM 471  
 QY 544 TEGRAG---MAAVASPTG-NCDLERFAQVLEKELPLYPARLPLPELHKTGTYKFKT 599  
 Db 472 IDEKAGEVPVAVVRNNGFTTTEEEKQFVSKQVYKIRIFRVFFVDAIPKSPSGKILRK 531  
 QY 600 ELR 602  
 Db 532 DLR 534

Search completed: July 16, 2001, 18:15:13

Job time: 189 sec

7